

Prize-winning multidisciplinary

The winner of a new prize exemplifies outstanding scientific breadth. But crossing boundaries brings challenges of its own, and progress in meeting them has been mixed.

The world is full of scientific prizes, provided by learned societies, industry, philanthropists and governments: some with big money attached, some tiny but equally prestigious, some that aim to elevate, others to ridicule.

It's welcome, therefore, that a new annual prize not only celebrates scientific excellence but can be bestowed for whatever additional positive criterion its judges choose. Earlier this year, the Kohn Foundation gave to the Royal Institution of Great Britain funds for an annual prize of £10,000, to be called the Henry Dale prize (named after the neuro-physiologist and Nobel laureate who directed the Royal Institution's laboratories). It is to be awarded annually to a life scientist who has worked in, or engaged with, the United Kingdom.

This year the committee (of which *Nature's* editor is a member) decided to invite nominations of scientists who, as the additional criterion, had made outstanding multidisciplinary contributions. Individuals who single-handedly embody multidisciplinary ideals are rare, after all.

As the entries showed, new techniques in physics, chemistry and mathematics, or old techniques applied in new ways to biological problems, provide opportunities for significant individual achievement. So too do combinations of clinical, epidemiological and basic biological research into conditions and disease.

It helps a candidate's case when the problem tackled is a fundamental one, when striking solutions have been found, and when they have applications in industry and implications for society. And above all it helps if the researcher has spanned the spectrum of scientific practice; has produced an interesting hypothesis that connects fundamental processes with evolutionary considerations and makes specific predictions; and then spends three decades pursuing its implications using mathematical modelling, experiment and evolutionary studies to demonstrate the power of that hypothesis.

This sums up of the career so far of this year's winner of the Henry Dale prize: Tom Kirkwood, head of the Department of Gerontology at the University of Newcastle upon Tyne. And as *Nature* exists above all to promote communication across the disciplines, it seems

appropriate to highlight such multidisciplinary science in action, as well as to point to its associated challenges.

It was a chance conversation in a lift with a researcher in ageing, Robin Holliday, that first set Kirkwood's mind on to the problem. A mathematician-turned-biologist, Kirkwood developed an idea that ageing is the result not of a pre-programmed set of biological events, but of an accumulation of damage that occurs because natural selection did not place a high priority on long-term cellular repair. According to this notion, evolution has delivered a trade-off in the use of resources between maintaining the quality of the organism and reproducing it — in favour of the latter. This is one of several current theories of ageing (see *Nature* **408**, 233; 2000), but since its publication in 1977 it has provided a rich seam of predictions and experimental tests, yielding insights of their own and leaving the hypothesis unscathed.

Inevitably, people studying a mechanism associated with ageing, such as telomere shortening or oxidative stress, focus on their own territory and may underestimate the importance of the combined effects of both. But Kirkwood has pursued modelling techniques that allow synergies between different influences to be investigated. He now looks forward to a greater emphasis on analysing both predictable and stochastic cellular behaviours of organisms in late age, made possible by advances in computation and experimental techniques.

Kirkwood knows as well as anyone the difficulties that can confront multidisciplinary research. Funding agencies now cope quite well in crossing disciplinary boundaries. But he has also seen difficulties in journals' peer-review processes when, as he puts it, a referee can see bridges being built from opposite sides of a disciplinary chasm but cannot judge where they should meet. He is not alone in his caution about institutions that draw together scientists from different disciplines in the hope that sparks will fly — mental proximity counts more than physical. But how to ensure that researchers working in more than one discipline, who have to be employed within one disciplinary department while collaborating with another, will not be seen as foreign by both when it comes to institutional recognition and promotion? This challenge is now more important to address than ever. ■

Committed, yes. And commitments?

Tony Blair's speech to the Royal Society last week was an important act of leadership. Now what?

Britain is a crucible for technological inventiveness and scientific creativity, and some of its industries know how to profit from it. Yet it has provided some dreadful examples of the interplay of science, media and the public in addressing controversial technologies. So it matters to sometimes-beleaguered researchers and technologists when its prime minister celebrates science as a key national activity and positions himself firmly in its favour against what he characterizes as the timidity that overshadows public debates. Such public angst is notorious enough, he said last week, for industrialists in Bangalore to tell him that they would leapfrog British technology because of it. (A full text of the speech is at www.nature.com/nature/blair.html)

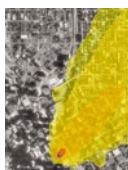
Tony Blair's speech may herald a renewed government campaign in

support of commercialization of genetically modified crops and animal experimentation, and comes ahead of the outcome of a review of public spending, due in the summer. Researchers should applaud Blair for his sustained support, and reiterate their wish-list for public funds, but with higher expectations. More money for science teachers and facilities in schools, rebuilding of university infrastructure, further improvements to pay and conditions for postdocs, ensuring that government and its agencies can commission independent research to sustain regulatory standards, and enhancing the resources available to independent agencies for providing public information and conducting public consultation in topics of significant social concern — such commitments are priorities if Blair's committed rhetoric is to yield substance. ■





Grounded
Cutbacks mean that Venus Express gets its wings clipped
p474



Defensive action
US bioterror bill receives all-round approval
p475



Plane sailing
Research on air turbulence cleared for take-off
p476



Bird watch
Zoos on the lookout for outbreaks of West Nile virus
p477

Priorities for genome sequencing leave macaques out in the cold

Erika Check, Washington

The chimpanzee, chicken and honeybee are among the top priorities for genome sequencing selected by the National Human Genome Research Institute (NHGRI) in Bethesda, Maryland. The announcement on 22 May sparked equal measures of jubilation and despair among researchers who had been lobbying the institute to fund projects for their favourite organism.

Other priorities include the sea urchin *Strongylocentrotus purpuratus* and the protozoan *Tetrahymena thermophila*. Fifteen fungi make up the remainder of the list.

There will be arguments among scientists about each selection, but the main point of controversy is the choice of the chimpanzee ahead of the rhesus macaque, which is more widely used in biomedical research but is listed as only a “moderate” priority.

The main NHGRI-funded sequencing centres — at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, Washington University School of Medicine in St Louis, Missouri, and Baylor College of Medicine in Houston, Texas — are fully occupied with the human, mouse and rat genomes. But as these projects wind down over the coming months, spare



Bright idea? Comparing the chimp and human genomes could shed light on human cognitive abilities.

sequencing capacity will open up, so the NHGRI asked researchers to submit ‘white papers’ making the case for the next round of organisms to be sequenced. An expert panel reviewed a list of 13 bids, ranging from the green turtle to the cow.

Individual sequencing projects will still

have to be approved one by one. But given the immense opportunities offered by genomics, advocates of particular model organisms had invested high hopes in getting onto the NHGRI’s priority list. “Working on a system that doesn’t have a sequence available is like doing research with one hand tied behind

Investigation into nanotechnology papers expands

Geoff Brumfiel, Washington

The probe into alleged data falsification by one of the brightest stars in nanotechnology is widening. Over the past week, the number of papers under investigation has more than doubled, as scientists scrutinize the publications of Jan Hendrik Schön of Bell Laboratories in Murray Hill, New Jersey.

Nanotechnology researchers were stunned when news broke that some of Schön’s papers on electronic devices made from organic molecules were being investigated. Graphs in separate papers describing the behaviour of different devices, operating under different conditions, contained plots that were

apparently identical (see *Nature* 417, 367–368; 2002).

On 16 May, when Bell Labs announced that a panel headed by physicist Malcolm Beasley of Stanford University in California would lead the investigation, five papers — three in *Science*, one in *Nature* and one in *Applied Physics Letters* — were under suspicion. Another *Science* paper was soon added to the list.

And with researchers submitting additional papers to the panel, the total had risen to 14 by *Nature*’s press deadline earlier this week. They include papers in *Physical Review B*, *Synthetic Metals*, *Thin Solid Films* and *Physica Status Solidi B*. Most were

published in the past two years, although two date from 1998.

“We will look into all substantive allegations,” says Beasley. But the panel will not examine Schön’s entire published output, which extends to more than 100 papers.

The researchers who found the first suspect figures say that they are now quitting the search. “I think the issues have been raised,” says Paul McEuen of Cornell University in Ithaca, New York. Lydia Sohn of Princeton University in New Jersey adds: “All of us are kind of burnt out. If people find more, I hope they will submit them to the panel.”

▶ your back,” observes Eric Green, director of the NHGRI’s own sequencing centre.

The priorities highlight animals useful in agriculture, such as the chicken and the honeybee, but not plants, for which sequencing projects are funded by the National Science Foundation and the US Department of Agriculture. The agriculture department is also expected to collaborate with the NHGRI in work on the chicken and honeybee genomes.

Fungi are included because they cause human and agricultural disease. Both the sea urchin and the chicken are model organisms for developmental biologists, and *Tetrahymena* is used in cell biology and genetics research, particularly for studies of the telomeres that cap the ends of chromosomes, which shorten as cells age. “The biologists who work on these organisms are really energized by this,” says Richard Gibbs, director of Baylor’s sequencing centre.

Members of the selection panel say they tried to select a range of organisms to allow evolutionary and comparative studies. Advocates for the chimpanzee, which is thought to have a genome 98.5% genetically identical to our own, argue that comparisons between the two sequences could yield clues about the origin of human cognitive abilities and language. They also say that the sequence could help to explain why chimpanzees do not get diseases such as breast cancer and Alzheimer’s — offering potentially important medical insights.

But these arguments cut little ice with researchers who have been lobbying for the rhesus macaque to be made the top-priority primate. Macaque advocates feel that the panel was swayed by the public appeal of sequencing our closest relative, and point out that scientists can do little work on chimpanzees because of welfare concerns and the animals’ endangered status.

“It is a tragedy that the NHGRI is committing these resources to sequencing the chimpanzee genome when there are so many investigators using rhesus monkeys who desperately need a sequence to advance research,” says John VandeBerg, director of the Southwest National Primate Research Center in San Antonio, Texas. He argues that rhesus macaques provide vital models in research on many human diseases. They are particularly important in attempts to develop vaccines against HIV, and are widely used in basic neuroscience studies.

But the macaque might still win favour in the future. The NHGRI’s priorities panel will review white papers three times a year from now on. And researchers can revise and resubmit their proposals, says Jane Peterson, programme director for large-scale sequencing at the NHGRI. “We don’t consider the high-priority bin to be static,” she says.

▶ www.nhgri.nih.gov/NEWS/sequencing.html

Delays and cutbacks kill off Europe’s mission to Venus

Quirin Schiermeier, Munich

There was good and bad news last week for Europe’s space scientists. Two projects believed to have been under threat — a mission to map star positions precisely and another to study the Sun — survived a meeting of European Space Agency (ESA) officials. But plans to send a spacecraft to Venus in 2005 were axed amid doubts over the project’s timetable.

The demise of the Venus Express mission came in Ardenes in arctic Norway, where ESA’s Scientific Programme Committee met to reassess the agency’s scientific activities for the next 10 years. ESA has to save up to 200 million euros (US\$184 million) during that period in response to budget cuts imposed last year by its members, which comprise 15 European states plus Canada (see *Nature* **415**, 730; 2002).

Venus Express was to have carried out the first global survey of Earth’s nearest planetary neighbour, including studies of its atmosphere and its interactions with the solar wind, the stream of charged particles emitted by the Sun. The mission had the potential to save money by using the same production facilities as Mars Express, an ESA mission that is scheduled to launch next year.

In the end, Venus Express was dropped because of diverging time schedules, says Giovanni Bignami, scientific director of the Italian Space Agency and outgoing vice-chair of the Scientific Programme Committee. The mission needed to be launched very shortly after Mars Express if the two were to share facilities, but funding for the spacecraft’s 20-million-euro payload of scientific instruments, which was to come from individual member states, had still not been secured.

The decision shocked those involved with the project, and has left planetary scientists feeling cheated. “It came out of the blue,” says Dimitri Titov, a physicist at the Max Planck Institute for Aeronomy in Katlenburg-Lindau, Germany, and science coordinator for the cancelled mission. “We are losing a huge opportunity to study an under-explored planet.”

“It is absolutely regrettable,” agrees Michael Grewing, director of the Institute for Millimetre Radiowave Astronomy in Grenoble, and chair of the group of space scientists that advises ESA. His panel had argued in favour of the project. “We have missed the unique opportunity of putting together a full-range planetary mission for just one-third of the usual costs.”

Grewing points out that Venus Express would have fitted well in the six-year gap



Missed chance: Venus Express was originally expected to share facilities with Mars Express.

ESA

between the launches of Mars Express and Bepi Colombo, the planned 2009 mission to Mercury. He says that ESA should now look for ways to help Solar System researchers, such as increasing European participation in planetary missions run by NASA.

Two projects that ESA officials had previously hinted could be axed were retained at the Norway meeting. GAIA, an advanced astrometry satellite that will precisely map the brightness, colour and relative positions of a billion objects in the sky, could be launched by 2009 if development of the spacecraft’s instruments can be completed in time. Space scientists hope that data from GAIA will improve their understanding of stars’ inner workings, and help them to detect possibly tens of thousands of extra-solar planets, including small, Earth-like ones. Solar Orbiter, which will study the Sun from a distance of around one-fifth of the Earth’s solar orbit, will be launched between 2008 and 2013 as planned.

GAIA

▶ astro.estec.esa.nl/GAIA

Solar Orbiter

▶ sci.esa.int/home/solarorbiter/index.cfm

UN predicts long wait to repair environment

Virginia Gewin

The benefits of some environmentally friendly policies will not be apparent until decades after they have been enacted. That is one of the messages of a report from the United Nations Environment Programme, which, even by the standards of global environment assessments, is sobering reading.

Global Environmental Outlook 3 (GEO-3), a study of the links between environmental, social and development issues, contains a range of dire but familiar predictions about the impact of factors such as climate change and industrial development. But the report, released last week in the run-up to August's World Summit on Sustainable Development in Johannesburg, was unusually pessimistic about the prospects for reversing the damage.

The new predictions are contained in one of four possible futures outlined in the report. The authors considered scenarios in which global politics were dominated by concerns over either markets, environmental and social policies, security or sustainability. These were based on attempts to quantify the effect of the different approaches on population levels, economics, technology and governance.

Some of the scenarios produced a famil-



iar picture. In a world dominated by a market mentality, for example, land and forest degradation becomes a critical issue, particularly in Latin America and the Caribbean.

But the sustainability scenario's predictions shocked some of the authors. "The delays between changing human behaviour and environmental recovery came as the biggest surprise to the regional experts," says Jan Bakkes of the National Institute of Public Health and the Environment in Bilthoven, the

Netherlands, one of the report's authors.

The report found that even if environmentally friendly approaches were adopted now, carbon dioxide concentrations would continue to rise until 2050. Water shortages would continue and coastal pollution would increase slightly. Bakkes blames difficulty in altering energy and transport infrastructures.

Originally used during the 1950s to simulate future conflicts, scenarios were revived in an improved form by the Intergovernmental Panel on Climate Change (IPCC) in the early 1990s. "By adding scenarios to assessments you come up with a credible story about how the world might evolve and can translate that into quantifiable information," says Bert Metz, also at the Bilthoven institute and co-chair of the IPCC working group on strategies for tackling climate change.

More than 1,000 scientists contributed to GEO-3, which divides the world into no less than 17 different regions. By contrast, the IPCC has used just four regions in previous assessments, although the panel's new chair, energy economist Rajendra Pachauri, has pledged to improve regional detail in future studies (see *Nature* 417, 106; 2002).

♦ www.unep.org/GEO/geo3

Scientists placated as US bill gets tough on bioterror

Virginia Gewin, Washington

Eight months after the anthrax attacks that brought the bioterrorist threat into sharp relief, the US Congress has passed a \$4.6-billion bill to prepare for future assaults. It includes measures to control researchers' access to microbes that can be used as weapons — but scientific societies say they

are happy with the new legislation.

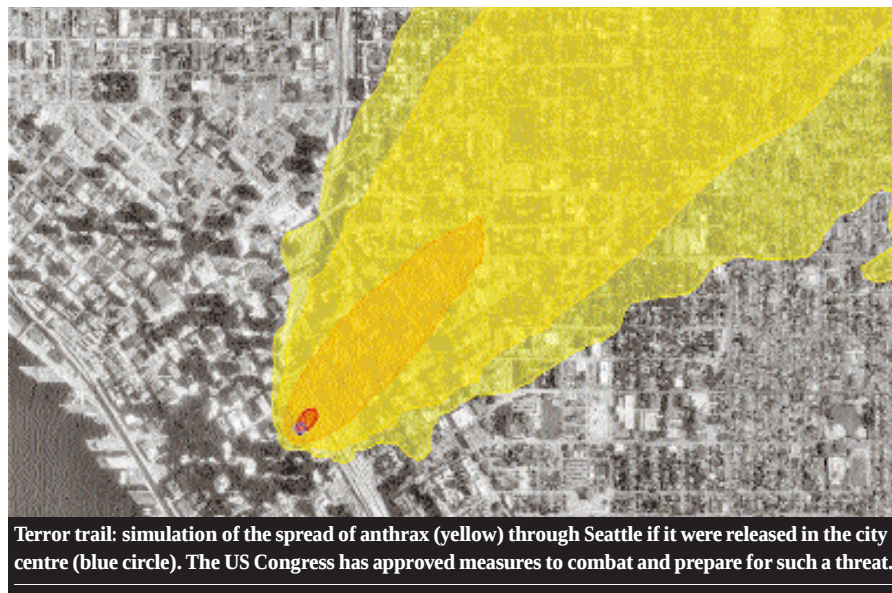
Under the bill, researchers from countries listed by the Department of State as sponsors of terrorism — Cuba, Iran, Iraq, Libya, North Korea, Sudan and Syria — will be banned from working on 42 microbes and toxins in US labs. Anyone with a criminal record will also be excluded, and security for

labs working with the agents will be stepped up. A national database will also be created to monitor research on these agents.

Last autumn, the House of Representatives backed a bill to prevent all non-citizens apart from permanent residents from working with the agents — which alarmed scientists who felt this would restrict legitimate research (see *Nature* 414, 3–4; 2001). But scientific societies are satisfied with the final legislation. "We feel pleased that Congress has worked with us to come up with a reasonable set of regulations," says Janet Shoemaker, director of public and scientific affairs for the American Society for Microbiology.

The bill also provides funds for stockpiling drugs and vaccines, and for improving drinking-water security. To protect the food supply, funding will be provided for research and training to detect, prevent and respond to agricultural bioterrorism. As part of this effort, an early-warning system will coordinate diagnostic labs, agricultural research facilities and public-health agencies.

The bill also addresses terrorism against animal experiments. "We are pleased that it includes increased criminal penalties for terrorist activities against animal facilities," says Jordan Cohen, president of the Association of American Medical Colleges. ■



Clouds lift for turbulence-detector initiative

Rex Dalton, San Diego

Plans to develop a system for detecting air turbulence are gearing up again, three years after a lawsuit brought by an airline pilot halted the original research.



Blue sky research: after a three-year hiatus, efforts to understand air turbulence are back on course.

Unexpected turbulence causes hundreds of injuries every year on commercial flights and has resulted in the deaths of at least three people. But atmospheric physicists have had limited success in understanding the phenomenon, and it is difficult to predict precisely when and where turbulence will occur.

In 1995, Larry Cornman, a physicist at the National Center for Atmospheric Research (NCAR) in Boulder, Colorado, published details of software that can detect and quantify turbulence, and which gathers data on the conditions that create it. The software monitors the plane's vertical acceleration along with a dozen other factors such as altitude, speed and air temperature. A measure of the turbulence can then be transmitted to other aircraft approaching the same area (L. B. Cornman, C. S. Morse & G. Cunniff, *J. Aircraft* **32**, 171–177; 1995).

A new version of Cornman's software is now being installed in aircraft. But the researchers involved say that a large portion of the US commercial fleet could already have been equipped with the system were it not for a lawsuit that halted the project in 1999.

The software had by then been installed in about 80 Boeing 737s as part of a collaboration between NCAR researchers, United

Airlines and aerospace company AlliedSignal (now merged with Honeywell, an engineering firm in Morristown, New Jersey). But the project stalled when Gary Bush, a pilot with Delta Airlines, filed a federal lawsuit in Fort Worth, Texas, alleging that the system infringed a patent he received in 1996 for a turbulence-detection system. Funding from the US Federal Aviation Administration (FAA) was withdrawn while the lawsuit was contested.

But NCAR researchers claim that the patent contains only vague details of a detection system and should not have been issued because of Cornman's earlier publication. Bush got no compensation when the case was settled last year, and the agreement left the NCAR free to use the data collected by its system for non-profit research.

Bush says that he never expected the FAA to stop funding the research, or the airlines to stop installing the software system, and that this is one of the reasons he settled the case.

The scientists who developed the software say that the case taught them an important lesson: don't rely on the published record of research to ensure free and unlimited access to a new technology. Securing a patent to ensure wide accessibility may be a better course of action, they say.

M. WAGNER/AVIATION-IMAGES.COM

Internecine rows threaten to sink whaling commission

David Cyranoski, Tokyo

The annual meeting of the International Whaling Commission (IWC) is usually a fractious affair. But the recriminations surrounding this year's event, held in the former Japanese whaling port of Shimonoseki last week, have left some observers fearing for the organization's future in the face of increasingly bitter stand-offs between pro- and anti-whaling lobbies.

In 1986, the IWC voted for a moratorium on commercial whaling to protect diminishing stocks. The commission's scientific committee has since devised a computer model to allow sustainable quotas to be set, but the majority of anti-whaling nations has upheld the ban.

Japan is lobbying to overturn the moratorium, citing survey results that show that some whales, such as the minke, are abundant enough to support an end to the ban. But Japan's research programme, which includes the killing of some 600 whales per year, has garnered criticism — not least because the meat often ends up being sold for human consumption in Japan.

Japan's plan to take a further 50 minke and 50 sei whales was attacked this year in a report submitted to the IWC's scientific committee by a research team led by Phil Clapham of the Northeast Fisheries Science

Center in Woods Hole, Massachusetts. The programme "provides no testable hypotheses and thus no reasonable criteria by which to judge its results," the report claims.

As expected, Japan's attempt to overturn the moratorium was rejected. But the big surprise was the voting down of a US–Russian request for a five-year extension to an agreement that allows aboriginal communities in Alaska and Chukotka, northeast Russia, to take a yearly quota of 280 endangered bowhead whales.

Japanese delegates, who voted against the extension, said that they opposed it because of US "hypocrisy" in rejecting a proposal for Japan's traditional whaling communities to be allowed to take 50 minke. The vote "showed the anti-whaling countries that they couldn't do whatever they want," says Seiji Ohsumi, director of the Institute of Cetacean Research in Tokyo. But if such tit-for-tat actions continue, there is a danger that the IWC will fall apart, raising the spectre of unregulated commercial whaling.

The IWC's scientific committee seems to be just as divided as its parent body. "It just puts together contradictory reports from the scientists representing the two sides," laments Toshio Kasuya, a whale-ecology researcher at Teikyo University of Science

and Technology, west of Tokyo, who sits on the committee as an independent scientist.

Last week, Japan's fisheries agency said that its scientists will not take part in a meeting next month to analyse the techniques used in the country's research, citing scheduling difficulties. But some scientists claim that Japan is unwilling to have its research critically evaluated.



All at sea: tit-for-tat action cast a shadow over this year's meeting of the whaling commission.

KYODO NEWS

US zoos keep watch for cross-species killer

Jonathan Knight, Salt Lake City

American zoos have begun monitoring for outbreaks of West Nile virus, which has killed thousands of birds and 18 people in the United States since 1999. It is the first time that zoos have been used for surveillance of a zoonotic disease — one that can pass from animals to people.

Current monitoring relies largely on reports of crow deaths from the public, but this approach may be less effective in the western United States, where the public is only dimly aware of the disease. "With wild animals you need a pretty big die-off before anyone notices, and you lose that window of opportunity for rapid response," says Tracey McNamara, a veterinary pathologist at Bronx Zoo in New York, who helped to develop the zoo-based monitoring scheme.

West Nile virus is endemic throughout Africa and parts of Europe, but had never been seen in the Americas before 1999. That summer, wild crows began dying in New York, but state epidemiologists were unable to determine the cause. The breakthrough came when McNamara dissected the brains of crows that dropped dead on her zoo's grounds and found lesions consistent with viral infection. By early September, a Guanay cormorant, three Chilean flamingos and a bald eagle had also died.

McNamara suspected a link to the simultaneous outbreak of human encephalitis in the area and confirmed that viral particles in the birds' brains were the same size as those from human patients. The Centers for Disease Control and Prevention (CDC) in Atlanta identified both viruses as West Nile.

State public-health agencies have tracked the disease mainly by examining dead crows and mosquitoes, which can spread the virus. Insecticides are used to control infected mosquito populations. But birds can infect one another directly, and the virus is still spreading through the United States. In 1999, it was found in only four states, but by the end of 2001 it had turned up in 27, reaching as far south as Florida and west to Iowa.

All zoos monitor their animals closely for illness, principally by taking regular blood samples. Participants in McNamara's programme now also send blood to Cornell University's veterinary diagnostic lab in Ithaca, New York, to be tested for antibodies against West Nile virus or for the virus itself.

Together with Dominic Travis, a veterinary epidemiologist at Lincoln Park Zoo in Chicago, McNamara started the programme in September with a network of six zoos. It now includes 64 zoos nationwide, with 20 more about to join. By March, some 4,500 samples had been submitted, of which more than 60 tested positive, McNamara told the annual meeting of the American Society for



In the pink: but West Nile virus has killed others from this Chilean flamingo flock at Bronx Zoo.

Microbiology in Salt Lake City last week. So far, the system has not identified an infection outside the known endemic area. But the real test is just beginning, as mosquitoes become more active in the summer.

The CDC has agreed to spend \$50,000 on the pilot project, which will cover the cost of

reagents, a technician and shipment of samples. If it proves successful, the scheme could be expanded to include regional testing centres for more rapid detection. "If it looks effective, we are eager to continue it," says the CDC's Jim Hughes, who directs the National Center for Infectious Diseases in Atlanta. ■

Superbugs reveal chink in armour

Jonathan Knight

Most antibiotic-resistant lung infections are caused by a handful of bacterial strains, microbiologists now believe, contrary to previous opinion. This homogeneity raises hopes for vaccines, the American Society for Microbiology heard last week in Salt Lake City.

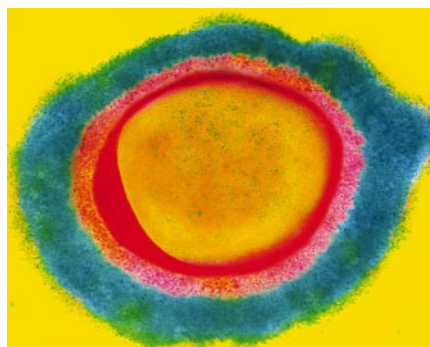
Bacteria can share antibiotic-resistance genes through the transfer of loops of DNA called plasmids. Although the staphylococci that cause hospital-acquired and childhood

pneumonia do not carry plasmids, they can incorporate plasmid DNA into their genomes. This was thought to explain the worldwide spread of scourges such as methicillin-resistant *Staphylococcus aureus* (MRSA).

It now seems that resistance is spreading not because diverse strains are acquiring plasmids, but through clonal reproduction of a few 'superbugs'. Microbiologist Keith Klugman of Emory University in Atlanta, Georgia, co-founder of the Pneumococcal Molecular Epidemiology Network, presented evidence that about 75% of drug-resistant childhood pneumonia cases are caused by 10 strains of pneumococcus. Almost half are caused by one strain, Spain 23-F.

Herminia de Lencastre, a molecular biologist at Rockefeller University in New York, said her laboratory tested 3,000 MRSA isolates from 14 countries and found that 70% of infections belonged to five strains.

Low natural variability makes resistant bacteria good targets for vaccines. Vaccines against MRSA are still under development, but the pneumococcal vaccine Prevnar is showing signs of success against antibiotic-resistant strains in South Africa. ■



Weak spot: researchers now believe that the MRSA bacterium has very little diversity.

US admits spraying sailors with sarin during Cold War tests

Washington Hundreds of US sailors were sprayed with deadly biological and chemical agents, including the nerve agent sarin, in secret Cold War tests, the US government admitted last week.

Papers released by the Pentagon on 23 May show that the sailors were exposed to the weapons between 1964 and 1969 in six experiments to test the navy's vulnerability to chemical and biological attacks. The papers were made available after enquiries by veterans who had been involved in the trials.

The overall plan for the experiments called for the sailors to be given protective clothing, but officials say it is unclear whether they were all wearing the clothing, or whether they all gave their consent to taking part.

British scientists involved in similar trials, in which one airman died after being exposed to sarin, face possible prosecution (see *Nature* **416**, 468; 2002).

♦ <http://deploymentlink.osd.mil>

Kyoto Protocol closer to fruition as Japan signs up

Tokyo The Kyoto Protocol, the international agreement on limiting greenhouse-gas emissions, has almost secured enough signatures to be enacted.

For the treaty to come into force, it must be ratified by at least 55 countries, which together contributed at least 55% of global carbon-dioxide emissions in 1990. Last week Japan's lower house of parliament passed the protocol, making the country the agreement's 55th signatory.

But even with 55 signatures, the protocol falls short of the emissions criterion. With the United States refusing to ratify the agreement, and Canada and Australia still debating it, reaching the 55% target requires a signature from Russia, which accounted for 17.4% of 1990 emissions. Experts believe that Russia will sign up, but suggest that it may delay formal ratification until later this year in an attempt to secure better terms.

Japan's politicians are now discussing a bill on new regulations that would enable it to meet its commitment to reduce emissions to 6% below those of 1990.

Hawking fights for ban on 'fraudulent' book

Washington Astrophysicist Stephen Hawking has asked the US Federal Trade Commission to recall a book of his lectures.

The Theory of Everything, published by New Millennium Press of Beverly Hills, California, is the transcript of a Cambridge



Stephen Hawking:
wants book withdrawn.

lecture series that Hawking gave in 1989. Hawking agreed to publication of a transcript when the lectures were recorded by Dove Books On Tape, but his lawyers say this was intended only to cover a transcript included with the audio cassettes — not publication of a separate book by New Millennium Press, set up by the former head of Dove.

The move constitutes "a fraud on the public", according to the complaint, because the lectures mirror the Cambridge University academic's 1988 bestseller *A Brief History of Time*. Trade-commission complaints are an unusual way to try to stop publication, but Hawking's representatives say the 60-year-old author, who suffers from motor neurone disease, is too weak to appear in a US court.

Oceanographers escape from stricken submarine

San Diego A fire on a US Navy research submarine, the *USS Dolphin*, forced crew and scientists to abandon ship in high seas shortly before midnight on 21 May amid fears that the vessel could sink.

The fire and flooding occurred about 160 kilometres southwest of San Diego, California, while the diesel-powered vessel was at the surface charging its batteries.

All 43 crew and scientists on board were saved by an accompanying research vessel, and suffered only minor injuries. They were evacuated by boat to an oceanographic ship that was in the area. A US Coast Guard helicopter rescued two crew members who fell into the water.

The *USS Dolphin* carries an array of oceanographic research equipment and has been used to perform deep-water studies and near-bottom surveys. The submarine was towed on 23 May into San Diego, where navy officials are investigating the incident.



Sub in trouble: the *USS Dolphin* was towed into port after fire and floods forced an evacuation.

Scientists speak out on sustainability

San Diego Natural and social scientists from more than 40 countries met at the National Autonomous University of Mexico in Mexico City last week in an attempt to shape the scientific agenda for August's World Summit on Sustainable Development in Johannesburg. The event was run by several organizations, including the Initiative on Science and Technology for Sustainability, a multidisciplinary network.

The meeting, attended by 350 researchers, completed a two-year effort to decide on a series of projects to recommend for debate in August. These include improvements to water and ecosystem management in Africa, prediction of extreme weather events, and aquaculture projects for Southeast Asia.

A full report is to be posted this week on the Forum on Science and Technology for Sustainability, hosted by Harvard University.

♦ <http://sustainabilityscience.org>

Rearrange letters to spell a new word in the story of life

London There's a new word in life's vocabulary. The letters of DNA can be rearranged to spell out a 22nd amino acid, researchers have discovered.

The scientists who cracked life's genetic code in the 1950s said it contains a mere 20 'words' — the amino acids from which life's myriad proteins are built. A 21st amino acid was discovered in bacteria in 1986. Now researchers at Ohio State University in Columbus have found another, pyrrolysine, in a microbe called *Methanosarcina*.

The microbe lives in the guts of cattle, where it breaks down organic molecules for energy, producing methane as a by-product. Pyrrolysine is in one of the organism's methane-producing proteins (G. Srinivasan *et al. Science* 296, 1459–1462; 2002 and B. Hao *et al. Science* 296, 1462–1466; 2002).

"We thought the 21st was an aberration, but here we see another one," says Michael Chan, a biochemist on the Ohio State University team. "Perhaps 23 and 24 are just around the corner."

Neutrality versus the niche

According to some ecologists, you don't need to invoke adaptation to explain biodiversity. They may sound like nihilists, but their ideas are proving remarkably resilient. John Whitfield reports.

Vive la différence? Forget it. *La différence est morte*. That is the message from the proponents of neutrality, a view of ecology that is anathema to many in the field. Rather than focusing on how differences between species allow them to coexist, neutrality assumes that trees in the rainforest, or corals on a tropical reef, are basically all the same.

Ecologists have strived to explain species' patterns of distribution, abundance and coexistence for more than half a century. The traditional explanation is that each species is adapted to exploit a unique niche — shady or sunny, wet or dry, and so on. But neutral theories assume that all organisms are equal, and consider only factors such as random dispersal, the birth and death of individuals and the total number of organisms in the community.

The troubling thing for most ecologists is that neutral simulations can produce ecosystems that look just like the real thing. "Neutrality starts with assumptions that are clearly wrong, but produces patterns that match what we see in nature," says Jonathan Levine of the UK Natural Environment Research Council's Centre for Population Biology at Silwood Park, west of London.

Sceptics say that recreating patterns isn't the same as understanding the mechanisms that cause them, and complain that the assumptions of neutrality are practically



Graham Bell (left) and Stephen Hubbell's neutral models can describe rainforests and reefs (top).



Testing time: tundra plants (above and right) are helping to reveal the limits of neutral models.

impossible to test. The debate isn't purely academic — our ideas of how biodiversity arises will influence how we attempt to conserve it. "We'd better figure these things out," says Stephen Hubbell of the University of Georgia, Athens, one of the gurus of neutrality.

Put it in neutral

Hubbell, a specialist in the ecology of tropical forest trees, last year published *The Unified Neutral Theory of Biodiversity and Biogeography*¹, which has become the bible of neutrality. Hubbell's neutral model considers only the birth and death of individuals, random dispersal, the overall population of individuals in the ecosystem, the total number of species and the origin of new species. The models favoured by neutral ecology's other leading proponent, Graham Bell of McGill University in Montreal, Canada, are even more austere, leaving out the birth of new species^{2,3}.

Neutral models consider the ecological properties of every individual in the population to be identical. Individuals compete for the same, fully exploited pot of resources, but the identity of the winner of any single contest — the tree that fills a gap in the canopy, for example — is left to chance.

Get the balance between birth, death and dispersal right, and you get a near-perfect recreation of natural communities. "It predicts

the individual species abundances of more than 800 tree species in a Malaysian rainforest with just three numbers," says Hubbell of his model¹. Neutral theories also make predictions about evolutionary history, for example that rare species will have arisen recently — although Hubbell acknowledges that these will be difficult to test. "By the time a species is detectable it's already really old and really abundant," he argues.

Hubbell admits to being taken aback by neutrality's success in mimicking the natural world. "I'm very puzzled by how well it's worked, and I was really unprepared for how theoretically rich in questions it was," he says.

If correct, neutral theory means that the species in a habitat have been thrown together — and will come and go — at random. This insight, Hubbell believes, might be useful in designing nature reserves. If species are closely adapted to fit an ecological niche, communities will be relatively stable, and hard to invade. Reserves can therefore be small. But if species are more equal, and come and go at random, communities will be more fluid, and bigger reserves will be needed to protect rare species from the buffeting of chance. "My view is that most communities are open and easily invaded," he says.

But if species are essentially all the same, does it matter if we lose one of them? That is a dangerous line of argument, says Bell. If just



L. M. STONE/NPL

1% of species perform a vital and specific ecological function, he explains, this will make little difference to the statistical predictions of neutral ecological models. But take those species away, and a previously healthy ecosystem will be in serious trouble.

Even the staunchest supporters of neutral ecology accept that the theory has limitations. It only applies within one level of the food web — it might explain the diversity of trees, for example, or the diversity of herbivorous insects, but not how the number of tree species might affect the number of herbivorous insect species. Hubbell says that the theory is more likely to hold for plants or microbes — where different species overlap more in the way that they exploit resources — than animals. Moreover, it breaks down on large spatial scales. To give an extreme example, alpine plants are clearly adapted to different conditions from those in which lianas and mahogany trees thrive, so would struggle to survive in the rainforest.

The big question is whether biodiversity really results from neutral ecological processes — indicating that ecologists have overestimated the importance of the niche — or whether the models' accuracy is a coincidence. Some ecologists, such as Levine, believe that neutral models are most useful as an ecological 'null hypothesis' for revealing the differences between their predictions and real ecosystems. Others point out that niche-based models can also produce a good

description of the natural world. This year, for instance, a team led by Jérôme Chave, now at the Laboratory of Terrestrial Ecology in Toulouse, part of the CNRS, France's national research agency, showed that both niche and neutral models can reproduce natural patterns of species abundance⁴.

A simple life

Hubbell counters that theories that stress the importance of the niche generally have to be much more complex than neutral models to achieve superior results. "To say that 100 or more parameters gives a better fit is not very satisfying as a critique," he says.

The niche may still be a valid concept, Hubbell suspects, albeit not as important as ecologists have long supposed. But Bell goes further, arguing that there is little experimental evidence to support the idea of niches. If plants are closely matched to their local environments, they should perform poorly if transplanted. But that isn't what Bell found when his team tried moving plants within Canadian forests⁵. "The results puzzled us — they clearly didn't point to any powerful degree of local adaptation, and sometimes the rarest species was the most successful in a new location," he says.

Bell is now looking at tundra plant communities to try and work out the scales on which the predictions of neutral theory hold. The key test, he says, is how likely species are to be present together. If particular species

tend to occur together, it would suggest that some environmental factor determines community composition, says Bell. But more haphazard patterns of cohabitation would support the neutral theory.

Scale bars

Other studies have produced mixed results. One, published earlier this year by a team including Chave and Hubbell, found that, in Panamanian and Amazonian forests, random dispersal could explain patterns of tree diversity between different areas at scales of between 0.2 and 50 square kilometres. However, the neutral theory broke down at smaller and larger scales⁶. A study of sawflies feeding on birch trees in Finland, which showed that species are specialized to feeding on leaves of a particular age⁷, has also been cited⁸ as evidence against the strongest interpretations of neutrality.

Some sceptics, meanwhile, still find it impossible to accept that neutral theories can yield real insight into ecological processes. They also point to the difficulty of validating Hubbell's model, given that it includes parameters such as the rate at which new species arise, which cannot readily be measured. Neutral models are intriguing, concludes Peter Petraitis of the University of Pennsylvania in Philadelphia, who studies rocky-shore communities. "But they don't offer much insight into what's operating underneath. It doesn't push us forward."

Bell senses the tide turning his way, however. "At first people thought neutral theory was nonsense. Now they think there's something in it," he says. Bell compares the debate to one that began more than 20 years ago among population geneticists about whether changes in gene frequencies are driven primarily by natural selection or by random 'genetic drift'. After initial resistance, most researchers now accept that genetic drift can be a significant factor. Bell suspects that it may take two decades for ecologists to reach a similar consensus.

While the debate goes on, even those who are unconvinced about neutral theory's validity value its stimulating effect on community ecology. "It's forcing us to address fundamental questions and work out what we really think," concludes Sean Nee, an evolutionary biologist and ecological geneticist at the University of Edinburgh, UK. ■

John Whitfield works in Nature's news syndication team.

1. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, New Jersey, 2001).
2. Bell, G. *Science* **293**, 2413–2418 (2001).
3. Bell, G., Lechowicz, M. J. & Waterway, M. J. in *Integrating Ecology and Evolution in a Spatial Context* (eds Silvertown, J. & Antonovics, J.) 117–138 (Blackwell Science, Oxford, 2001).
4. Chave, J., Muller-Landau, H. C. & Levin, S. A. *Am. Nat.* **159**, 1–23 (2002).
5. Bell, G., Lechowicz, M. J. & Waterway, M. J. *J. Ecol.* **88**, 67–87 (2000).
6. Condit, R. et al. *Science* **295**, 666–669 (2002).
7. Kause, A. et al. *Evolution* **55**, 1992–2001 (2001).
8. Agrawal, A. A. & Van Zandt, P. A. *Trends Ecol. Evol.* **17**, 62 (2002).

To catch a wave

A new observatory is about to search for the ripples in space-time that emanate from the Universe's most violent events. But despite its huge price tag, the detector might not spot anything. Geoff Brumfiel finds out why.

On the muddy floodplain of central Louisiana, two concrete tubes four kilometres long cut a giant 'L' into an expansive logging plantation. Inside the tubes, laser light bounces back and forth between mirrors, creating a measuring device so sensitive it can pick up the rumble of traffic entering the site, and the vibrations as nearby trees are felled.

But as evening falls and the logging stops, preparations continue for a far more ambitious piece of detective work. By the end of this year, the Louisiana detector will be looking for gravity waves — the faint ripples in space and time generated by colliding black holes and exploding stars. If detected, the ripples will provide new tests of Einstein's theories of gravity, and a way of peering into unseen areas of the Universe.

But that's a big if. The crashing trees and passing trucks — not to mention minor earthquakes — are hampering the final development of the detector. And because theoretical physicists are not sure what the waves look like, no one knows exactly what to expect. Leaders of the project — the Laser Interferometer Gravitational-Wave Observatory (LIGO) — admit that the Louisiana device, and its companion at Hanford in Washington state, might not initially spot anything. But, given that the detectors have a total cost of US\$296 million, some researchers are wondering how they ever got built.

No one has ever detected a gravity wave directly since they were first postulated 90 years ago — although physicists concur that



Bent on success: the L-shaped LIGO facility in the middle of Louisiana (above) hopes to corner gravity waves produced by events such as a black hole colliding with a neutron star (simulated, top).

the waves do exist. Albert Einstein's general theory of relativity, published in 1916, predicts that large masses in rotating systems should produce ripples in space-time (see 'General relativity for beginners', opposite). Einstein realized that these gravity waves would stretch and squeeze any object that they pass through, but dismissed the idea that anyone could ever detect them.

A typical gravity wave passing through the Earth, for example, would stretch the planet by just 10^{-16} metres. "Einstein thought they were just too difficult to detect," says Rainer Weiss, a physicist at the Massachusetts Institute of Technology.



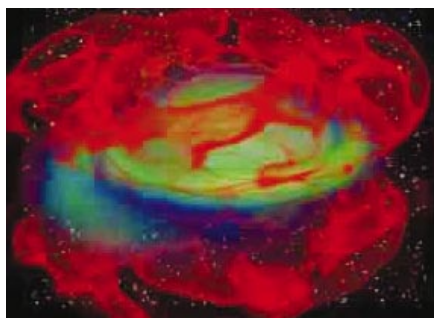
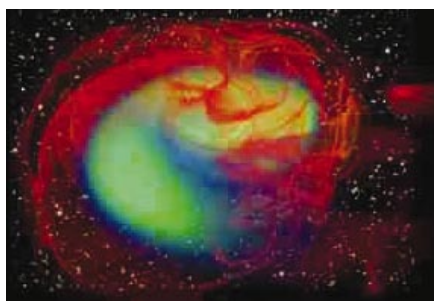
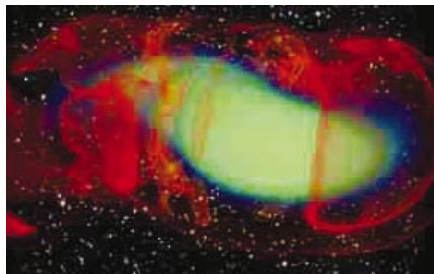
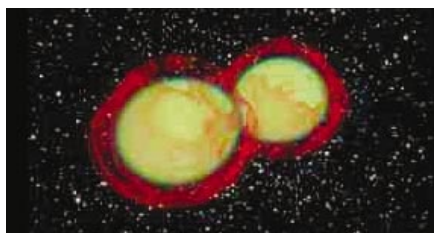
The first gravity-wave hunter, Joseph Weber.

The chance to detect gravity waves directly is exciting because it

would confirm Einstein's theoretical prediction. But the extreme events that create the most powerful waves provide a further incentive. "It's a completely different world," says Weiss. "If you're near a black hole, space is so contorted and curved that straight lines don't extend very far. Nearby clocks move at different rates, and nothing stands still." And, because dust and debris often surround these objects, many believe that the only way to learn about them is through the gravity waves some configurations of black holes emit.

Attempts to detect the waves date back to the 1960s. The first detector, the brainchild of Joseph Weber, a physicist at the University of Maryland, College Park, consisted of an aluminium bar 2 metres long and 1.5 metres in diameter. Weber predicted that the bar would ring as if struck by a hammer if a passing gravity wave caused its length to expand and contract momentarily.

In the late 1960s, Weber stunned the



Early insight: spiralling neutron stars, modelled here, offered the first evidence for gravity waves.

physics world by announcing that he had detected gravity waves passing almost simultaneously through bar detectors in Maryland and Chicago. The waves appeared to have changed the length of the bars by 10^{-15} m. But sceptics pointed out that waves this powerful could only come from events involving a significant fraction of the Milky Way's mass. "If you actually calculated what he was allegedly seeing, the Galaxy would be wiped out in about a million years," says Weiss. As time went by, it became clear that Weber's statistical analysis of his results had been flawed.

But Weber's announcement captured the imagination of many young scientists. "I was fascinated by it," recalls Tony Tyson, a physicist at Bell Laboratories in Murray Hill, New Jersey, who was a graduate student at the University of Chicago at the time. By the mid-1970s, Tyson had built larger bar detectors, as had other groups around the world. Although none successfully detected gravity

waves, interest in the field continued to grow, thanks in part to the first indirect evidence for the existence of the waves.

In 1974, physicists Russell Hulse and Joseph Taylor, then at the University of Massachusetts in Amherst, used a radio telescope to observe a pair of neutron stars orbiting one another. Hulse and Taylor realized that the stars would be emitting gravity waves, and so would be slowly losing energy and moving closer together. They tracked the stars for four years, and in 1978 announced that their orbits were changing in exactly the way predicted by Einstein's theories. The finding, still the best observational evidence for gravity waves, helped win Hulse and Taylor the 1993 Nobel Prize in Physics.

Pulling together

Now the race was on to build a device that could detect gravity waves directly. "We knew the scientific pay-off from success would be tremendous," says Kip Thorne, a theoretical astrophysicist at the California Institute of Technology (Caltech) in Pasadena. Thorne and others at Caltech were keen to host the device, and in the late 1970s decided to recruit Ronald Drever, an experimental physicist from the University of Glasgow, UK, to help them design a new kind of detector.

Drever had experience of building L-shaped devices called Michelson interferometers. Incoming laser light is split into two beams at the corner of the L, and each beam races down one leg to a mirror where it is reflected back to centre. As the beams recombine, they create an interference pattern with a dark spot over a light-sensitive detector. If a gravity wave, or any other disturbance, changes the length of either leg, the intensity of light at the detector changes. By studying the change in intensity, researchers can determine the change in the light's path. Interferometers are not necessarily any more sensitive than Weber bars, but they can detect a much wider range of frequencies of gravity wave.

The Caltech team was joined by Weiss,

who had been working independently on interferometer detectors. In 1985 they submitted the proposal for LIGO to the National Science Foundation (NSF), calling for a pair of interferometers, which they claimed would detect disturbances of about 10^{-19} m in the relative path lengths of the lasers. Hundreds of millions of dollars would be needed to construct their devices. Yet right from the start, the researchers were not completely certain whether they would detect anything.

The scientific uncertainties clouding LIGO were, and still are, twofold. On one hand, the theorists were unable to predict the precise size and frequency of the waves that they hoped LIGO would detect, raising doubts about the optimal detector design. And no one was sure how regularly gravity waves reach Earth. If collisions between black holes are rare, for example, it might take centuries to spot a single event.

There were also technical problems. Researchers working on Weber bars experienced minor earthquakes that constantly shook their labs. The tremors posed a significant problem, as their frequencies — between 0 and 100 hertz — are similar to those of the gravity waves that theory predicts should be most abundant.

Designed for life

To avoid these difficulties, the NSF advised that the project should have the potential to be upgraded if the initial design failed to spot anything. "We were told many times by the NSF: 'Don't go for something that is just going to end after one attempt,'" says Weiss.

The plan was to start with two geographically separate facilities large enough to hold high-power lasers, seismically isolated optics, and several interferometers. Using two detectors would allow potential sightings to be double-checked.

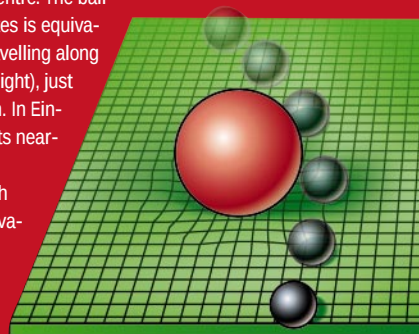
The first interferometers were to be built from commercially available components. Their main function would be to test the sophisticated electronic and computing systems needed to run the experiment. The

General relativity for beginners

Albert Einstein's general theory of relativity describes gravity as a distortion in the four dimensions of space and time. Visualizing this is difficult, but a two-dimensional analogy makes it clearer.

Imagine a sheet of rubber with a bowling ball at its centre. The ball is a planet, and deformation of the rubber surface it creates is equivalent to the way that real planets distort space. Objects travelling along the sheet will roll down the slope towards the bowl (see right), just as gravity would draw a passing comet towards the Earth. In Einstein's theory, the Earth attracts objects because it distorts nearby space, just as the bowl distorts the rubber sheet.

Now imagine two bowling balls spiralling around each other. This would create ripples on the rubber sheet, equivalent to the gravity waves produced when two stars orbit each other. By studying the ripples, observers can learn about the stars' movements.



team would then install more sensitive second-generation interferometers.

In 1989, the NSF gave the plan its formal backing. "There was a comfort factor with LIGO that you don't always get with big-money projects," says Frederick Bernthal, the NSF's deputy director during the latter half of LIGO's planning period. "Even if you never saw a gravity wave, the leading-edge technology involved with building LIGO was just extraordinarily impressive."

Political persuasion

The LIGO team now had to convince Congress that the project was worth funding. Robbie Vogt, the Caltech physicist who was then director of LIGO, led an intensive year-long lobbying campaign. One of his speeches caught the attention of Bob Livingston and Bennett Johnson, then a Louisiana representative and senator, respectively. They were impressed by the idea, and wanted to use the cutting-edge science — and the associated jobs — to boost the economy of their state.

The NSF helped to smooth the way by giving the Livingston site in Louisiana a high rating in a survey of possible locations. And at the time, no one thought to worry about the disturbances caused by falling trees. The site for the other detector at Hanford was on unwanted, government-owned land that had previously been used for nuclear-weapons construction.

Although it had the NSF's blessing, some scientists remained worried about the project's cost. Many were, and still are, reluctant to criticize publicly a project that has the backing of prominent universities and the government, but some voiced their concerns at the time. "As a physicist I am fascinated by this experiment and would like to see much of it funded," Tyson told the science committee of the House of Representatives in March 1991, "but not at the expense of hundreds of individual principal investigator grants." As part of his testimony, Tyson submitted comments from more than 60 physicists and astronomers, all expressing doubts about the plans. But LIGO was already on a roll, and had support from the Louisiana politicians. Congress approved the project in autumn 1991.

With the political battle won, researchers have spent the past decade designing and building the two observatories. Both are now almost ready to begin the search for gravity waves. But the start of data acquisition will not be marked with a fanfare — hopes of immediately detecting gravity waves remain low.

Barry Barish, a high-energy physicist at Caltech and the current director of LIGO, says that whether or not you expect LIGO to see anything still depends on what model of gravity-wave production you favour. "And the theories are kind of crummy," he adds. In the years since LIGO was first proposed, estimates of the frequency and power of detectable events have actually decreased.



The construction of LIGO has inspired other similar projects, such as the GEO 600 in Germany.

"They were wildly optimistic in their initial assessments of the potential sources," says Jerry Ostriker, a theoretical astrophysicist at Princeton University in New Jersey. Ostriker, a long-time critic of LIGO, believes that today's estimates give even the future versions of the giant machines only a slim chance of seeing anything.

Noise, noise, noise

The technical difficulties have also persisted. Construction of the facilities finished over two years ago. But in 2001, a 6.8-magnitude earthquake at the Hanford site knocked mirrors and other optics against their mounts and delayed the project by three months. The most difficult issue, however, has been filtering out one source of noise after another. The tree-felling at Livingston restricts operations to evenings and weekends. Even at these times, LIGO must take account of phenomena such as 'Earth tides' — the shifting of the Earth's core every 12 hours caused by the Moon's gravitational pull.

These effects can be compensated for by electronically adjusting the mirrors to filter out false signals. But each time the researchers filter out a signal, they must be sure that they have not interfered with previous adjustments. It is a bit like tuning in a faint radio signal over the static, explains Mark Coles, director of LIGO's Livingston observatory. "It's sort of unglamorous," he adds.

Some researchers question whether the noise will ever be reduced to a level that will allow gravity waves to be detected. "There's this big question mark: how much further is noise going to allow them to go?" says Tyson. Data collection, which had been due to start this summer, has now been put back until the end of the year at both sites.

But LIGO researchers remain confident. "It wouldn't be a total surprise if we detect something fairly early," says Barish. "But I'd be shocked if we don't within a decade." And after three years of operation, the researchers

plan to apply for money to upgrade LIGO. They hope, for example, to add more powerful lasers and to improve seismic isolation.

But whatever the results, LIGO will have boosted the field of gravity-wave detection. Since it was commissioned, similar projects have begun in Germany, Italy, Australia and Japan. One of these — Virgo, based near Pisa in Italy — will probably come online after LIGO in 2003 and should be similarly sensitive. Doubts remain about whether these other projects can overcome the problems, such as seismic noise, faced by LIGO, but if more than one of the devices detects waves, they could pool their data to pinpoint the source.

Plans for a space-based gravity-wave detector are also in the works. Physicists in Europe and the United States are designing an interferometer that would travel around the Sun in an orbit similar to that of the Earth. Known as the Laser Interferometer Space Antenna (LISA), it would consist of three satellites in a triangle, each separated by 5 million kilometres. LISA would have a similar sensitivity to LIGO, but should find it easier to capture the elusive low-frequency signals that seismic noise can obscure. If funding can be secured, work on LISA will begin within a few years, with the aim of launching the spacecraft by the end of the decade.

Projects such as LISA show that LIGO is more than just a detector — its construction has also helped to build a community of researchers. In the United States, the giant devices have become nuclei for hundreds of gravity-wave researchers at 30 institutions across the country. "We have experimenters who are looking at the next generation of detectors. We have theorists who are interested in the data," says Weiss. "And they're not old like me. That's the biggest development."

Geoff Brumfiel is *Nature's* physical sciences correspondent in Washington.

LIGO ♦ www.ligo.caltech.edu

Virgo ♦ www.virgo.infn.it

LISA ♦ lisa.jpl.nasa.gov

How affluence harms the environment in Europe

Increased use of fertilizer is not an option in crowded lands where it is already a problem.

Sir—Stephen Budiansky has discussed in *Correspondence*¹ how, in the United States, affluence could be good for the environment. He argued that although the US population grew by 90% and the real per capita disposable personal income grew by 177% between 1945 and 1997, the per capita area used for food production, urban areas and other land uses remained unchanged, or even declined.

The per capita land use for food production in the United States (1.69 ha) is slightly lower than in developing countries (sub-Saharan Africa is on average 2 ha), whereas food production is several magnitudes higher. Budiansky cites higher use of fertilizer as the main factor accounting for the higher yields per hectare. Two new reports^{2,3} give us the opportunity to translate this view into a western European perspective.

Flanders in Belgium is among the richest and most developed regions in the world. Between 1990 and 2000 the population grew by 3%, while the real per capita disposable personal income grew by 60%. In contrast to the United States, the areas for food production and habitation

have grown by 5.5% and 24%, respectively, over the past ten years. Compared with the United States (agriculture 1.69 ha; urban and other land use 0.13 ha), land use per capita in Flanders is very low (agriculture 0.14 ha; urban and other land use 0.06 ha). However, because of the high population density, there is only 0.23 ha per capita of Flemish territory available (2.8 ha in the United States). Therefore, about 62% of Flanders is allotted to food production, while 27% is covered by urban areas and other land use, leaving only about 11% for nature (about 30% in the United States).

Budiansky does not discuss the environmental impact that goes beyond land use (see ref. 4 for one example). Agriculture is extremely intensive in Flanders, emitting Europe's highest levels of nutrients into the environment. The OECD nutrient balance database for 2001 (see www.oecd.org) shows that during the past decade the nitrogen surplus (from manure) was between 177 kg and 203 kg per ha of total agricultural land. The critical load for grasslands in nature reserves and natural grasslands in agricultural areas was exceeded in all sampled points². As a consequence,

vegetation of nutrient-rich environments gradually replaced that of nutrient-poor environments throughout Flanders³. Over-fertilization is causing species extinction; for example it is one of the main reasons why nearly a third of the area's butterfly species have been wiped out during the past century⁵. The rate of extinctions is speeding up, with half the remaining species expected to die out during this century.

Although present policy aims to tackle the problem at the source by limiting nutrient emissions, atmospheric nitrogen deposition is barely decreasing, and nutrients will continue to accumulate for some time to come³.

Luc De Bruyn*†, Jan Scheirs†, Hans Van Gosssum†

*Institute of Nature Conservation, Kliniekstraat 25, 1070 Brussels, Belgium

†Department of Biology, University of Antwerp (RUC), Groenenborgerlaan 171, 2020 Antwerp, Belgium

1. Budiansky, S. *Nature* **416**, 581 (2002).

2. Van Steertegeem, M. *MIRA-T 2001* (VMM, Garant, 2001)

3. Kuijken, E., et al. *Nature Report 2001 in a Nutshell* (Institute of Nature Conservation, Brussels, 2002).

4. Reay, D. *Nature* **417**, 15 (2002).

5. Maes, D., Van Dyck, H. *Biol. Conserv.* **99**, 263–276 (2001).

O no: opossums are not at risk in Tasmania

Sir—In your welcome News story highlighting the severe threat to Tasmanian wildlife by the introduction of the red fox from the Australian mainland¹, your claim that farmers and foresters use sodium monofluoroacetate (1080) to control opossums and wallabies that graze on crops and native trees in Tasmania is not correct.

First, opossums do not occur in any part of Australasia, including Tasmania; they are found only in North and South America, where they are carnivorous marsupials (polyprotodonts). Possums occur in Australia and are herbivorous marsupials (diprotodonts). The perception that opossums live in Australia dates back to the Cook and Banks voyage of 1770, when an arboreal marsupial seen in Queensland was described as being similar to the Brazilian opossum.

It is more than 20 years since the Australian Mammal Society adopted a list of common names for Australian mammals. This included a recommendation that the 'o' be dropped from the name of these Australian arboreal herbivores².

Second, the possums and wallabies that are the target of 1080 poison graze grass but browse trees. Browsers feed mostly on twigs and leaves from trees and shrubs, such as the eucalypt seedlings often planted by foresters and farmers in Tasmania, whereas grazers eat mostly grasses and herbs³.

The main point of your story — that the red fox severely threatens Tasmania's wild life, and that effective action is required to counteract this now — is undoubtedly true. Your News story states that up to 77 known species are under threat; we can be sure that future generations will look back to see how well their politicians and biologists served us at this time.

Thomas L. Montague*,

Andrew T. D. Bennett†

*Roe Koh & Associates P/L, PO Box 789, Mornington, Victoria 3931, Australia

†School of Biological Sciences, University of Bristol, Bristol BS8 1UG, UK

1. Dennis, C. *Nature* **416**, 357 (2002).

2. Kerle, J. A. *Possums* (UNSW Press, Sydney, 2001).

3. Strahan, R. *The Complete Book of Australian Mammals* (Angus & Robertson, Sydney, 1988).

Possums were correctly identified by the author of the original story. The error was inadvertently introduced during editing — Editor, Correspondence

What can Venter teach students about ethics?

Sir—In your Opinion article "Selling science to the young" (*Nature* **417**, 1; 2002) outlining what should be done to attract young people back to science, you remark that "many graduates do turn their science degrees into decent financial as well as spiritual reward".

I was shocked to read, on the very next page, the News report (*Nature* **417**, 3; 2002) about Craig Venter, who, after making a fortune, now, as your headline puts it, "lays foundations for fresh career as ethical philanthropist". In his own words, he plans to "spend some time examining my own genetic code".

What kind of image of science does this create? Let's hope that there still exists a pool of young scientists with a passion and vocation for doing basic research, otherwise cash-strapped universities and government research organizations will soon run out of personnel. It's a sad day for science if the criterion of success is measured in purely financial terms.

Rod Ceredig

Institut National Scientifique et de la Recherche Medicale, 17 rue des Martyrs, F-38054 Grenoble, France

When the Earth moved

How Earth scientists developed the theory of plate tectonics.

Plate Tectonics: An Insider's History of the Modern Theory of the Earth

edited by Naomi Oreskes

Westview Press: 2001. 448 pp. £24.99, US\$35, Can\$52.95

Jason Phipps Morgan

Editor Naomi Oreskes gave each of the 17 contributors to *Plate Tectonics* an interesting assignment. She not only asked them to write about their experiences as participants in the 'mobilist Earth' revolution in Earth science during the 1960s, when it was discovered that the continents are passive riders on the handful of tectonic plates that form the top boundary layer to Earth's convecting mantle, but also to offer their thoughts on what we can learn from this experience about the processes of scientific discovery. In her preface, Oreskes presents an interesting synthesis of what she feels we can learn from these tales of "science working well". She also provides a concise introduction to the state of Earth science at the beginning of the 1960s — in particular, the reasons why the scientific precursor to plate tectonics, continental drift, was firmly rejected in the 1930s to the 1950s.

The result is a fascinating book that has much to entertain and teach practising scientists, science managers, science funding managers, and anyone interested either in Earth science or in tales of scientific discovery. One of its best qualities is its multiple-perspective view on the eternal tension in science between ideas and observations. Some would argue that science is driven by data. The explosive increase in our knowledge of the ocean basins during the 1950s (largely driven by the US Navy's desire to learn more about this arena for submarine warfare) was certainly key to the proposal and acceptance of the ideas of seafloor spreading and the Vine–Matthews–Morley hypothesis for the origin of the magnetic 'stripes' covering the ocean floor.

The latter idea was independently proposed by Fred Vine and Drum Matthews (published in *Nature*) and by Lawrence Morley (in an earlier submission rejected by both *Nature* and the *Journal of Geophysical Research* for being too outrageous). It linked the then radical speculation that the Earth's magnetic pole randomly reverses direction about every million years with the similarly radical hypothesis of seafloor spreading, in which the sea floor is created by quasi-continuous volcanic eruptions at a narrow, globe-circling, volcanic spreading centre. As the eruptions cool, they 'freeze-in' the current direction of the Earth's magnetic



Field work: this map of magnetic 'stripes' on the sea bed contributed to the theory of plate tectonics.

field, leading to the formation of alternating 'normal' and 'reversed' magnetic stripes on the sea floor that provide a magnetic record of the age of seafloor formation. This idea, once accepted, provided the key magnetic evidence to convince most scientists of the reality of seafloor spreading, a major step towards plate tectonics.

Yet relatively few of the major conceptual breakthroughs were made by the scientists most active in collecting the key marine data, almost all of whom worked at a few major oceanographic institutions. These scientists were certainly capable of doing this, yet they didn't. Perhaps data collection

demanding so much intellectual effort and attention that they couldn't spend the necessary extra time to synthesize the results and implications of their new observations — with the notable exceptions of Xavier Le Pichon, Lynn Sykes and Bill Menard.

Instead, a surprising number of the conceptual advances were made by a relatively small number of scientists working at a remove from data collection, at the universities of Cambridge and Princeton, under the scientific influence of Teddy Bullard, Tuzo Wilson and Harry Hess. For example, what set Jason Morgan (Princeton) onto his 1967 discovery of the Morgan–McKenzie–Parker

hypothesis of plate tectonics was Bill Menard's observation (published in *Science* in January 1967) that the great fracture zones in the Pacific were not great circles, as he had stated in his 1964 book *Marine Geology of the Pacific* (McGraw-Hill) — the last great synthesis of marine geology before plate tectonics — but instead were... he didn't know what. Because Morgan was familiar with the map projection used by Menard to demonstrate that the great Pacific fracture zones were not great circles, when he read this paper he immediately realized that they looked like small circles about a single common pole of rotation. This was his breakthrough moment. It must have helped that at the time he was sharing an office with fellow Princeton postdoc Fred Vine.

Plate Tectonics should also help to debunk a common 'creation myth' that is told in many science textbooks, in which a few discoverers have great ideas and the scientific community immediately falls into their enthusiastic trial and acceptance. Actually, the most common initial response is likely to be neglect. When Morgan gave his April 1967 talk to the American Geophysical Union introducing plate tectonics, it wasn't in a vacuum but at the end of a stirring session (chaired by Menard) in which other researchers had been presenting many new results on seafloor spreading. Yet of those who heard it, only Le Pichon grasped enough from the talk to drop what he was doing to work to test the idea and its possible implications. Le Pichon mentions how many of his colleagues returned from this session also highly stimulated, but to try to refine and improve their marine magnetic observations, rather than spend time testing a strange theory. Likewise, ideas are rarely accepted on the strength of a single paper, no matter how good it is. Only when enough trailblazers have mined a new idea does it begin to be accepted by the broader community.

Although 35 years have now passed, and several of the contributed tales do smooth and polish the rough edges of this science, a surprising amount of the truly gritty, uncertain and fun nature of scientific research shines through many of these essays. Oreskes has assembled a book that will sit at the head of the short list of books — including *Noah's Flood* by William Ryan and Walter Pitman (Simon & Schuster, 1999), John McPhee's *Annals of the Former World* (Farrar Strauss Giroux, 1998) and *Evolutionary Catastrophes* by Vincent Courtillot (Cambridge University Press, 1999) — that I would strongly recommend to my fellow researchers, as supplementary reading to students taking their introductory courses in Earth science, and to friends interested in a good scientific yarn. ■

Jason Phipps Morgan is director of the Geodynamics Department at the GEOMAR Research Center for Marine Geosciences, Wischhofstrasse 1-3, Geb. 8D, D-24148 Kiel, Germany.



Word play: studies of language acquisition can help to find new ways of teaching children to read.

With brains, minds and voices

Foundations of Language: Brain, Meaning, Grammar, Evolution

by Ray Jackendoff

Oxford University Press: 2002. 498 pp. £25

John C. Marshall

The study of linguistics has made significant headway over the past 50 years. But that progress has often been accompanied by ferocious (and often personal) rows that have typically produced more heat than light. In *Foundations of Language*, Ray Jackendoff is scrupulously fair to any serious suggestion and incorporates relevant advances from all sides of the argument. Consequently, he will, no doubt, be attacked from all sides, but this should not obscure the fact that his new book is a masterpiece.

Here are just a few of the hitherto entrenched (and interrelated) positions that Jackendoff either cuts across or endeavours to integrate. First, the idea that the capacity to acquire language is underpinned by a biological organ unique to *Homo sapiens*, versus the notion that children learn to play language games as they would learn any other social skill. Second, that combinatorial syntax is the core of the language faculty, and the view that effective communication is all (and you can forget the dative case, progressive aspect and subject-verb agreement). Third, the idea that the formal structure of language can be sensibly studied as an abstract 'mental' object, versus the claim that only the neural networks underlying language abilities have any kind of relevant 'scientific reality'.

More importantly, Jackendoff works out in some detail how all three basic levels of any spoken language — phonology (the structure of sounds, syllables and metres), syntax (the structure of sentences) and semantics (the meanings of words, phrases and sentences) — are generative. That is, larger

units are built out of smaller ones by rule-governed iterative processes. Finally, correspondence (or interface) rules connect these three levels of description in what Jackendoff calls a 'parallel competence model', in which all levels have an equivalent combinatorial status. For example, "if the semantic structure of a sentence contains an Agent (the entity bringing an action about), it (normally) corresponds to the subject in syntactic structure".

Thus far, Jackendoff is characterizing what a language is and what kinds of implicit knowledge a person must have to be an English (or Spanish or Turkish) speaker. But unusually for a linguist, he also discusses how that competence can be realized in performance — in speaking and understanding speech. One issue here concerns which "aspects of an utterance must be stored in long-term memory and which aspects can be constructed online in working memory". This is a topic that has been discussed practically non-stop since the study of language acquisition was revived in the 1950s: a child can infer the regular rules for forming the past tense in English (for example, walk/walked and jump/jumped), and can accordingly construct such forms when they are needed. Yet many verbs, such as 'go', are exceptions to the rule. Thus in any utterance that requires the past tense of 'go', 'went' must be retrieved from long-term memory.

This much is uncontroversial (except for a few exponents of the wilder forms of associationist psychology). But then Jackendoff argues fairly convincingly that idioms and other only semiproductive constructions play quite a large role in English (and presumably in other languages); items that must be stored are more pervasive than is often acknowledged. Consideration of how different types of memory are deployed in language production then leads Jackendoff to a sketchy but provocative processing model "whose components correspond directly to the components in the parallel model" of linguistic structure and where

processing procedures are “isomorphic to the rule types in the parallel grammar”.

This notion of a relatively direct relationship between the competence model of structure and a performance model of language production and perception was first put forward in the 1960s (the derivational theory of linguistic complexity), with little success. Jackendoff's newer version has processes specific to particular linguistic levels interacting on a blackboard-like working memory that is dedicated to language.

These conjectures sound reasonable, although they cannot be properly evaluated until the details have been fleshed out much more explicitly. Similarly, Jackendoff's “evolutionary perspective” on the phylogensis of language draws plausible conclusions from the successes and failures of teaching chimpanzees an approximation to American Sign Language, and from Derek Bickerton's work on pidgin and creole languages. But once again the gulf between speculation and hard evidence is immense.

The concluding chapters of *Foundations of Language* concern meaning and reference. These tightly argued sections provide a superb and in many ways novel introduction to lexical and phrasal semantics, and to

the relationship between language and the world. On the latter topic, Jackendoff, contrary to the prevailing philosophical winds, attempts “to explain how reference and truth are attached to linguistic expressions” in a world “as conceptualized by the language user”. These chapters merit close study — and the book as a whole deserves a wide readership. ■

John C. Marshall is in the Neuropsychology Unit, University Department of Clinical Neurology, Radcliffe Infirmary, Oxford OX2 6HE, UK.

Views of a scientific American

The Age of Science: What Scientists Learned in the Twentieth Century

by Gerard Piel

Perseus Press: 2001. 460 pp. £24.99

Zaheer Baber

Gerard Piel recalls that his interest in science was ignited by his sociology tutor at Harvard University, Robert K. Merton, who was to

become one of America's most distinguished sociologists. This fortuitous intellectual encounter contributed to Piel's lifelong passionate relationship with science and his goal of making esoteric scientific discoveries accessible to the general public.

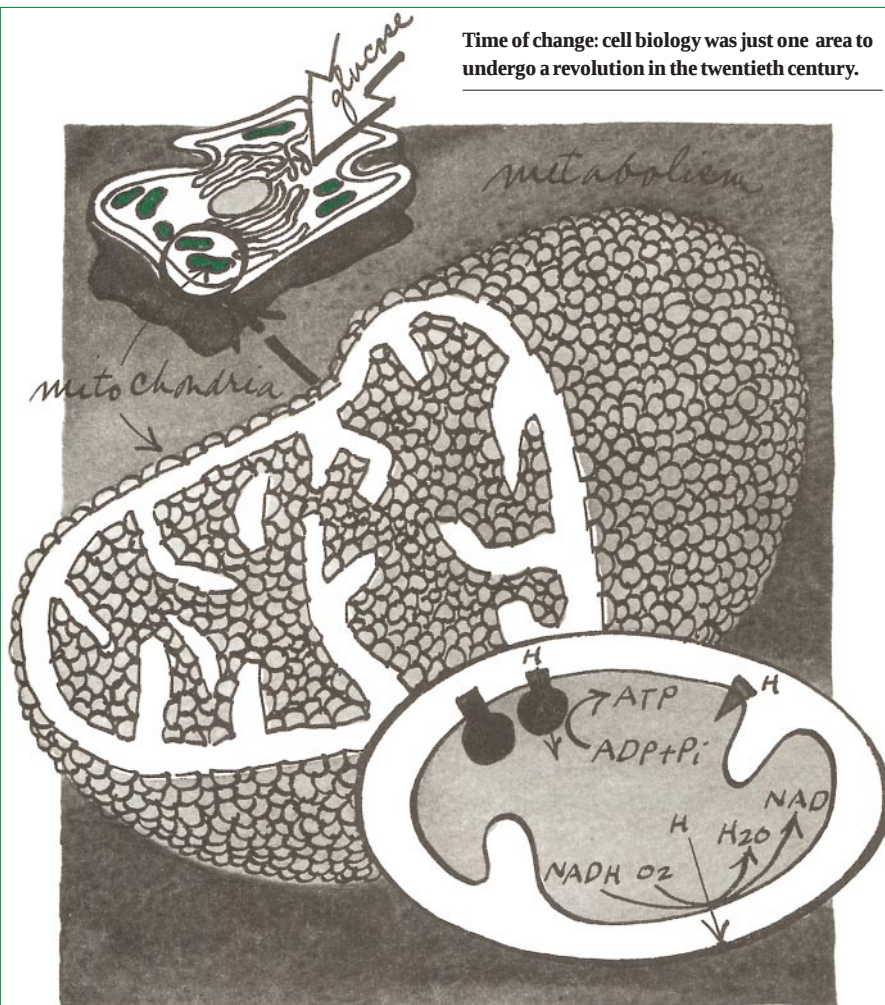
A brief stint at *Life* magazine, where Piel was assigned to cover science, led him to realize that there was little good coverage of scientific issues in US magazines or newspapers. So he took over *Scientific American*, a step that institutionalized his relationship with the leading scientists around the world. Now, after four decades of formal association with *Scientific American*, a term as president of the American Association for the Advancement of Science, and active involvement with several scientific and educational institutions, Piel is in a unique position to narrate the story of science in the twentieth century.

A brief account of the main intellectual developments across history sets the stage for this amazingly lucid mapping of the contours of scientific knowledge in the twentieth century. Special and general relativity, quantum electrodynamics, the fruitful collaboration between particle physicists and cosmologists, and the dramatic breakthroughs in genetics and evolutionary biology are just some of the major topics discussed by Piel. He incorporates them in a narrative that demonstrates not only his intimate knowledge of complex issues that cut across a very diverse variety of fields, but more significantly takes advantage of his formidable talent for making difficult concepts accessible to the general reader. The dozens of illustrations accompanying the text help to achieve the main objective of the book: enhancing public understanding of science.

Reflecting Piel's enduring commitment to civic engagement and interest in the sociology of science, he includes the changing social context and the personalities of scientists associated with the intellectual developments as integral elements of the narrative. Considering the changing pattern of funding and its implications for the production of scientific knowledge, he writes about “market-motivated industrial financing”, which contributes to a situation in which whole university departments “operate as subsidiaries of pharmaceutical companies” under agreements that “induce or compel restraint on publication and the open communication that is the life of competitive collaboration in science”. He says that this has reached the stage at which the question to be investigated “turns on the movement of NASDAQ exchange”.

On this issue, Piel's conclusion, developed towards the end of the book, is a pointed critique of the market, which he says has a dominant role in the deployment of specific technologies but has neither any

Time of change: cell biology was just one area to undergo a revolution in the twentieth century.



mechanism to internalize the cost of its operations to the environment nor any interest in responding to human need, purpose or hope. To address the inextricable intersection of science and technology with the social condition of humanity, he believes that we need to call on institutions other than the market.

Along the way, Piel discusses the impact of the Cold War arms race on the trajectory of scientific knowledge, the scientific actors behind the Manhattan Project to build the first nuclear weapons, the passions unleashed over the cancellation of the Superconducting Supercollider in 1994, the general unease over the entrenched commercial interests in the development of genomics, and debates about global warming and the carrying capacity of the planet.

The fact that Piel consistently connects the growth of scientific knowledge to the social and political milieu of its genesis, while simultaneously alluding to the changes in the social context induced by these discoveries, makes *The Age of Science* one of the best introductions I've seen to the world of science and its practitioners. I think it is easily the most intelligent, lucid and socio-logically sensitive discussion of the broad trajectory of scientific knowledge yet available. For anyone wishing to look beyond the 'science wars' to get the big picture of science, this book is the place to begin. ■

Zaheer Baber is in the Department of Sociology, University of Saskatchewan, Saskatoon, Saskatchewan S7N 5A5, Canada.

Pretty as a table: Gregor Mendel's meteorological data were presented largely without pictures.

The image shows an open book with two pages of handwritten meteorological data. The pages are filled with dense, organized tables of numbers and text. The title 'Meteorologische Beobachtungen' is visible at the top of the right page. The data is presented in a grid-like format, with columns and rows of numbers and text, organized in a systematic manner. The handwriting is in cursive, and the paper appears aged.

Science in culture

Peas without pictures

Gregor Mendel and the mathematical birth of modern genetics

Martin Kemp

Natural history before Gregor Mendel and Charles Darwin was full of visual delights. The publishing of books on the sciences of nature had been sustained in large measure by attractive prints of familiar and exotic specimens of flora and fauna, accompanied by texts which fluently combined observation, learned citation, philology, hearsay and legend. The Book of Nature, which stood alongside the Books of Scripture as providing us with access to the divine scheme of things, was lavishly illustrated. What, then, are we to make of the chastely non-visual publications of the two figures who were to rewrite the natural sciences, Darwin and Mendel? Darwin almost seems to have tested how far he could go before resorting to a picture, and Mendel's key discovery was expressible in ratios, algebraic letters and tables, rather than through images of plants.

There is a paradox at work here. Darwin and Mendel were supreme observers, relying on systematic handling of information obtained through rigorous scrutiny of things they had seen. Mendel's experiments with peas depended on the visual identification and methodical counting of traits that were significant in tracking the as yet unknown units of inheritance that passed through successive generations. Unrelenting visual and manual discipline was required for his meticulous pollination and data-gathering in the garden and greenhouse of the Abbey of St Thomas in Brno. Yet his crucial paper of 1865, published obscurely the following year in the proceedings of his local society for natural science, is devoid of illustrative interest. His results are laid out in rows of figures, with sets of

paired characteristics designated by capital and lower-case letters – Aa, Bb – in which the capitals signal the 'dominant characters'.

There is no single explanation for this shift away from the visual image in cutting-edge biology. There was clearly a general sense that if natural history was to be regarded as a 'hard' science, it needed to cast off its image as an attractive pursuit for amateurs and collectors of nature's wonders. In this, Darwin's impersonal theory – the disinterested mechanism of natural selection – combined conceptual abstraction and physical process in a way that seemed to promise the abandoning of subjectivity. Mendel was, in a way, even more radical. He resorted to mathematics, the province in the sciences of nature that had been the prerogative of physics and, to a lesser degree, chemistry.

What has become apparent in the work that I have done with Marina Wallace and our company Artakt for our exhibition about Mendel, "The Genius of Genetics", which opened earlier this month at the abbey in Brno, is the extent to which Mendel's studies in Vienna and the greater part of his activity as an investigator of natural phenomena were concerned with quantification in sciences other than natural history. His training in Vienna included studying with Christian Doppler, the first professor of experimental physics in Austria, noted astronomer and formulator of the Doppler effect. Mendel came to biology with a mastery of combinatorial mathematics, and his own publications centred on meteorology, which had become the most statistical of all the sciences of natural phenomena.

The library in the abbey still contains Mendel's annotated copy of August Kunze von Lichton's *Lehrbuch der Metereologie, Leichtfasslich Dargestellt* of 1850, an important textbook by the physicist on the taxonomy of clouds and the mathematical science of the atmosphere. The abbot's assiduous contribution to local meteorological records is evident in the sheets of patient and systematic data-gathering in the exhibition. Seen in this light, his genetic researches can be seen to belong with the nineteenth-century obsession with categorizing and quantifying as much as with traditional natural history and the practice of selective breeding in agriculture and horticulture.

The old cliché tells us that an illustration is worth a thousand words. For Mendel and those who gainsaid the seduction of pictures, thousands of acts of counting were worth more than either. Martin Kemp is in the Department of the History of Art, University of Oxford, Oxford OX1 2BE, UK.

▶ <http://www.mendel-museum.org>

The Genius of Genetics, A Celebration of Gregor Mendel Through Science and Art runs at the Abbey of St Thomas in Brno, Czech Republic until May 2003.

Two faces of water

David Chandler

Water and oil famously don't mix: the term hydrophobic (water-fearing) is commonly used to describe substances that, like oil, do not mix with water. Although it may look as if water repels oil, in reality the separation of oil and water in ambient conditions is not due to repulsion between water and oil molecules, but to particularly favourable hydrogen bonding between water molecules. Each water molecule can participate in four such bonds, sharing its two hydrogen atoms with two neighbouring water molecules and sharing two further hydrogen atoms associated with two other neighbours. Ice is a tetrahedrally ordered array, and liquid water a disordered network, of such hydrogen-bonded molecules.

Oil and water molecules actually attract each other, but not nearly as strongly as water attracts itself. Mixing enough oil with water therefore leads to a reduction in favourable bonding. Strong mutual attractions between water molecules induce segregation of oil from water and result in an effective oil–oil attraction, in the same way that groups of people segregate when one subgroup prefers to associate with its own kind. This water-induced attraction between oil molecules is called the hydrophobic interaction.

In the 1950s, Walter Kauzmann identified hydrophobic interactions as a primary source of protein stability. Proteins are chains of amino acids that fold into specific, functional, three-dimensional structures that are determined by the order of the amino-acid sequence. Kauzmann reasoned that as amino acids are either water-like or

oil-like, a given linear sequence would fold into the structure that best segregates oil-like amino acids from water. Later, when protein structures were deduced with the help of X-ray crystallography, Kauzmann's general predictions were found to be correct. Yet subsequent attempts to quantify his ideas by identifying a single parameter or function that characterizes the strength of hydrophobic interactions have been unsuccessful.

This difficulty has to do with length scales. Under the right conditions, hydrophobicity of the traditional Kauzmann sort can appear, but only for large oily surfaces in water. A different, weaker hydrophobic interaction occurs for small assemblies of oil molecules. Because the mutual attraction between water molecules is so strong, water–water interactions persist even in the presence of small oily species, perhaps a single alkane chain with, say, 10 or fewer carbon atoms. Hydrogen bonding simply goes around hydrophobic species. In the close vicinity of the oily molecules, the possible configurations of hydrogen bonding may be restricted, but the overall amount of hydrogen bonding remains relatively unchanged. Thus, the cost of hydrating a small, hydrophobic solute has more to do with the number of ways in which hydrogen bonds can form than with their strength. That is, the solvation free energy of the system is largely entropic and not enthalpic. However, this geometric picture breaks down for an extended oily region, because not all hydrogen bonds can persist near to its surface (see picture). The nature of hydrophobicity therefore changes when the size of oily surfaces depletes the number of hydrogen bonds. This energetic effect — the loss of hydrogen bonding — drives the segregation of oil from water.

What molecule size characterizes this crossover between small and large hydrophobic solute hydration? For a small solute where hydrogen bonding occurs adjacent to it, the number of water molecules that are affected by the solute is proportional to solute volume, and hence the solvation free energy is also proportional to this volume. The solvation free energies of ethane, propane and other similar molecules indicate that the proportionality constant is roughly 1×10^8 joules per cubic metre.

In contrast, the free energy of a large solute with an extended hydrophobic surface is proportional to the surface area of the interface between the solute and water. Here, the proportionality constant can be estimated from the surface tension of oil–water interfaces — about 6×10^{-2} joules per square metre. Thus, when the assembly interface extends beyond 1 nanometre, the energetic cost of assembling hydrophobic units is significantly more

Hydrophobicity

Oil and water molecules actually attract each other, but not nearly as strongly as water attracts itself.

favourable than the entropic cost of keeping them separate. The hydrophobic stabilization of smaller assemblies, however, is insufficient to overcome thermal agitation.

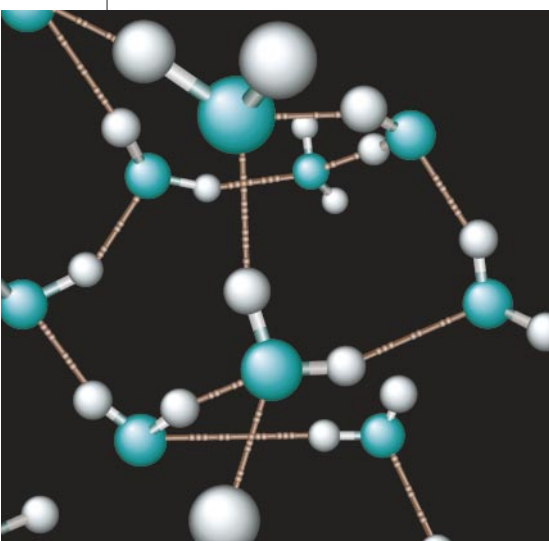
The crossover to assembly of hydrophobic units involves the nucleation of an interface with broken hydrogen bonds, such as that between water and its vapour. The assembly of hydrophobic structures requires the removal of water from regions between these groups; this is the same as vaporization. Hence, the closer water is to the liquid–vapour phase transition, the stronger is the tendency for hydrophobic assembly. I believe this explains why proteins are denatured by cold and by high pressure: lowering temperature and raising pressure both move water away from its liquid–vapour phase transition. Furthermore, raising the temperature increases the solvation free energy of small, disassembled hydrophobic units because hydration costs are entropic for such solutes, but it reduces the solvation free energy of larger, oily clusters because the costs here are enthalpic. Finally, in aqueous solutions with no assemblies of 1 nanometre or more in size, such as mixtures of water and simple alcohols, hydrogen bonding persists much as it does in pure water.

Without an aqueous environment, hydrophobic interactions cannot occur. But the free energetic cost of dissolving large, oily molecules in water is formidable. This cost can be successfully compensated for, however, by attaching sufficient numbers of polar and charged groups. In this way, large, oily molecules can easily mix with water, and nature can then exploit the phenomenon responsible for phase separation to create local assemblies of oily structures. These assemblies can exist on scales down to nanometres (but no smaller), because this threshold size coincides with the critical radius at which oil separates from water. This seems to explain the existence of cooperativity during protein folding, an idea that is worthy of experimental study. ■

David Chandler is in the Department of Chemistry, University of California, Berkeley, California 94720, USA.

FURTHER READING

Kauzmann, W. *Adv. Prot. Chem.* **14**, 1–63 (1959).
Stillinger, F. H. *J. Solut. Chem.* **2**, 141–158 (1973).
Lum, K., Chandler, D. & Weeks, J. D. *J. Phys. Chem. B* **103**, 4570–4577 (1999).
Ball, P. *H₂O: A Biography of Water* (Phoenix, London, 1999).
Tanford, C. & Reynolds, J. *Nature's Robots: A History of Proteins* (Oxford Univ. Press, 2001).



Closing ranks: water's strong hydrogen-bonding network (thin lines) can exclude other species.

Making it with molecules

Peter Zoller

Following the creation of atomic Bose–Einstein condensates in the mid-1990s, a major goal has been to produce a condensate of molecules. Technical challenges remain, but that achievement is now tantalizingly close.

Some atoms can undergo a phase transition at ultralow temperatures to form a Bose–Einstein condensate (BEC) — an extraordinary quantum phase of matter, in which all atoms pile up in the same quantum state. The heavier the atom species, the lower the temperature at which the condensation occurs. So can heavy particles, such as molecules, still be condensed? Moreover, molecules are composite objects with complex internal structure, which raises a serious conceptual question about whether a stable molecular condensate could ever exist.

For the past few years, creating a molecular condensate has been one of the main targets in BEC physics. If such a condensate were made, it might reveal new molecular spectroscopy and collision physics, and could open the door to new applications in ultracold quantum chemistry, or even quantum computing¹. Now, on page 529 of this issue, Donley *et al.*² report an experiment that is tantalizingly close to demonstrating the creation of a molecular condensate.

Since the first realization of a BEC of alkali atoms in 1995 (reviewed in ref. 3), we have enjoyed a feast of experimental demonstrations of the fundamental properties of BECs, from the coherence of atom lasers to superfluid behaviour. Loosely speaking, a condensate is to atoms what laser light is to photons: the remarkable coherence properties and brightness of laser light have their counterpart in the properties of atomic condensates. In their experiment, Donley *et al.*² have detected coherent oscillations between pairs of rubidium atoms and rubidium molecules in an atomic BEC. This is just a small step short of the final proof that the molecules formed in the atomic BEC have the coherence properties expected of a molecular condensate — a proof that will no doubt be given once the technical tools for direct imaging of molecular clouds are in place⁴.

There are several routes to creating a molecular condensate. One could follow the recipe for producing an atomic BEC³: molecules would first be precooled, then loaded into a magnetic trap⁵ where further cooling drives the gas below the critical temperature of the BEC phase transition. But the laser-cooling techniques developed in atomic physics do not necessarily work for molecules. So, clearly, new ideas are needed. For example, to reach ultracold temperatures a promising approach might be to cool molec-

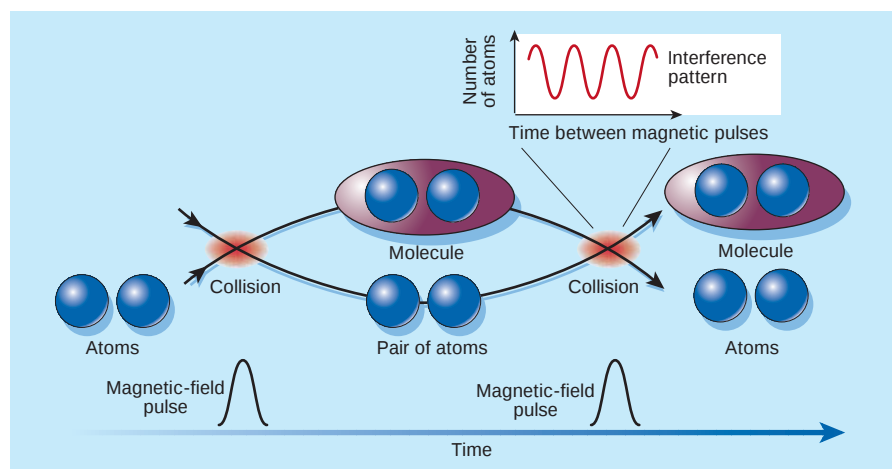


Figure 1 Towards a molecular Bose–Einstein condensate (BEC). Using a magnetic-field pulse to tune the energy of collision between rubidium atoms in a BEC, Donley *et al.*² created a superposed state of Rb atom pairs and Rb molecules. In the instant that a second magnetic pulse was applied, and the superposition broken, the authors measured the number of atoms in the condensate. A series of measurements taken with different time intervals between the two pulses formed a wave pattern typical of interference (upper right), proving that the atom and molecule states were quantum-mechanically superposed and not just classically mixed. This demonstration is a significant step towards creating a molecular, rather than atomic, BEC.

ules using an atomic BEC as a ‘refrigerator’.

Another approach — the one followed by Donley *et al.*² — is to start with an atomic condensate and then convert the condensed atoms into molecules. The key is the collisional interaction between the cold atoms. As two atoms approach each other and collide, laser light can be used to ‘capture’ the atoms during the collision and transform them into a molecule — a technique that has been demonstrated by Heinzen and colleagues^{6,7}. In their experiment with rubidium atoms, Donley *et al.* used a magnetic field to ‘tune’ the energy of the two colliding atoms so that the pair formed a ‘quasi-bound’, long-lived molecular state.

In general, however, only a fraction of the atoms will be converted into molecules. Thus, according to quantum mechanics, the system is in a ‘superposition state’ of atoms and molecules — that is, a state in which the particles are simultaneously atoms and molecules, rather than the simple mixture of atoms and molecules that one would expect classically. To distinguish the superposition from the mixture, Donley *et al.* exploit the quantum-mechanical interference between the superposed, but chemically different, atom and molecule states (Fig. 1). The setup is similar to a classic interferometer, where the molecules

represent one ‘arm’ of the interferometer and the atoms the other. Actually, as outlined in recent theoretical work⁷, this picture is somewhat oversimplified because thermal atoms produced in the collisions, and not only the atomic BEC, play an important part. Nevertheless, the interference patterns observed in this remarkable experiment provide the first evidence of a coherent quantum superposition formed between a pair of atoms and the molecular state.

Given that the creation of a molecular BEC is on the horizon, what could we expect to learn? First, there is much to be discovered about cold molecular spectroscopy⁴, and the collision properties of molecules are largely *terra incognita*. But using molecules also leads to new forms of condensates that are not encountered in atomic physics: molecules can have electromagnetic dipole moments and thus represent dipolar gases with rich varieties of shapes and phases⁸. These interactions also form the basis of the proposal by DeMille¹ to entangle molecules in a controlled way, for use in quantum computing. Molecular BECs could also be the starting point for ultracold chemistry in a pure quantum regime, revealing the physics of zero-temperature chemical reactions.

We are entering a new phase. The early

years of BEC physics were devoted to studying basic properties of condensates with the tools of atomic physics and quantum optics³, but molecular condensation is just one of the interdisciplinary developments that are leading us in new directions. Another is work with optical lattices⁹, in which dilute atomic gases (the starting point for atomic BECs) form strongly correlated systems with controllable interaction parameters¹⁰. A laboratory is now open to explore what has so far been only theory in condensed-matter physics. ■

Peter Zoller is on leave from the University of Innsbruck, Austria, at the Institut für Quantenoptik, University of Hannover,

Welfengarten 1, D-30167 Hannover, Germany.

e-mail: peter.zoller@uibk.ac.at

1. DeMille, D. *Phys. Rev. Lett.* **88**, 067901 (2002).
2. Donley, E. A., Claussen, N. R., Thompson, S. T. & Wieman, C. E. *Nature* **417**, 529–533 (2002).
3. Anglin, J. R. & Ketterle, W. *Nature* **416**, 211–218 (2002).
4. Masnou-Suuws, F. & Pillet, P. *Adv. Atomic Mol. Opt. Phys.* **47**, 53–127 (2001).
5. Weinstein, J. D., deCarvalho, R., Guillet, T., Friedrich, B. & Doyle, J. M. *Nature* **395**, 148–150 (1998).
6. Wynar, R. H., Freeland, R. S., Han, D. J., Ryu, C. & Heinzen, D. J. *Science* **287**, 1016–1019 (2000).
7. Kokkelmans, S. J. J. M. F. & Holland, M. J. Preprint cond-mat/0204504 (2002); xxx.lanl.gov
8. Santos, L., Shlyapnikov, G. V., Zoller, P. & Lewenstein, M. *Phys. Rev. Lett.* **85**, 1791–1794 (2000).
9. Greiner, M., Mandel, O., Esslinger, T., Hänsch, T. W. & Bloch, I. *Nature* **414**, 522–525 (2001).
10. Jaksch, D., Bruder, C., Cirac, J. I., Gardiner, C. W. & Zoller, P. *Phys. Rev. Lett.* **81**, 3108–3111 (1998).

Cell motility

Sharks' teeth and dunes

Laura M. Machesky

In order to move, cells need to push out protrusions known as lamellipods. These can vary greatly in their shapes, dynamics and contributions to motility, depending on their underlying molecular architecture.

What does it take to make a cell move? You might think that after more than 50 years of study, we could form a reasonable answer to this question. But the field of cell motility remains active and exciting, with models constantly being challenged and revised to accommodate new discoveries. For instance, Bear and colleagues¹ have just revealed at least two fundamental new insights into how cells move and change shape. As the authors describe in *Cell*, so-called Mena/VASP proteins help to determine how filaments made up of the actin protein are arranged at the front of a moving cell, to regulate the behaviour of the protrusions assembled there. Moreover, different types of protrusion seem to contribute differently to cell translocation, possibly permitting specialized types of cell motility in response to varied environmental cues.

One point that is emphasized by these results¹ is that protrusion of the front ('leading edge') of a cell does not always result in cell motility—although protrusions are certainly necessary for movement. Most models agree that a major force driving the advance of a cell's plasma membrane is the polymerization of actin into filaments. Cells are thought first to send out filopods, composed of thin parallel bundles of actin filaments that push out the plasma membrane in spikes. If conditions are favourable, a lamellipod—a thin, sheet-like protrusion with no detectable cytoplasm—can fill in around the filopods. The lamellipod can then be stabilized into a thicker protrusion, which usually points in the direction of movement. Most cells need to make lamellipods if they are to move. But the existence of more lamel-

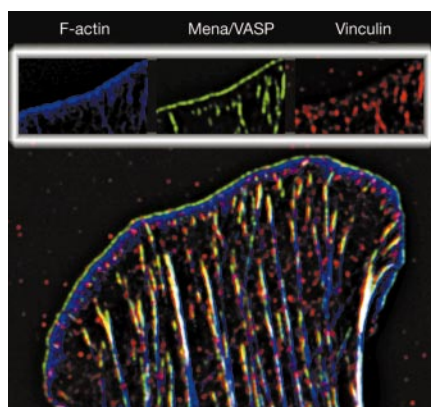


Figure 1 Molecular architecture of motile cells. The leading edge of a migrating mouse embryo fibroblast cell is shown, with F-actin (filamentous actin) stained in blue, Mena/VASP in green, and vinculin in red. Mena/VASP and vinculin co-localize (yellow) in focal adhesions—macromolecular complexes, important for cell movement, that link a cell's cytoskeleton to its extracellular matrix. Actin forms a fine band just behind Mena/VASP at the leading edge, as well as in stripes, known as actin stress fibres, that terminate in focal adhesion complexes.

lipods does not correlate with faster movement. Many filopods and lamellipods simply retract, and the cell recycles their actin for use elsewhere.

Indeed, when Bear *et al.*¹ studied cells that express Mena/VASP proteins, they observed lamellipods—which they described as sharks' teeth—that formed rapidly and frequently. But these were less productive in terms of their contribution to motility than the slowly rising and falling dune-shaped

lamellipods seen in cells lacking Mena/VASP. These observations help to resolve a controversy: Mena/VASP proteins can induce actin assembly *in vitro*^{2,3}, and localize to the leading edges of motile cells in quantities proportional to the rate of protrusion⁴; yet cells that lack Mena/VASP proteins move more quickly than cells expressing them⁵. The results of Bear *et al.* indicate that, without Mena/VASP, the actin structures that cells assemble might be more stable and contribute more efficiently to movement than those in the highly dynamic, but slower-moving, Mena/VASP-containing cells.

So how do the sharks' teeth and dunes compare in terms of their molecular architecture? Using an electron microscope, the authors found that the actin filaments in these two types of protrusion had very different structures. The slow-moving dunes had highly branched networks of filaments, pointing at angles to the plasma membrane. In contrast, the Mena/VASP-rich sharks' teeth contained long, parallel actin filaments pointing directly towards the plasma membrane. Presumably, the structures in dune-shaped lamellipods are assembled through activation of the Arp2/3 complex, which causes branches to sprout from the sides of pre-existing actin filaments⁶.

One implication of these results is that highly branched networks, such as those assembled by the Arp2/3 complex, are stable, slow to advance and slow to retract. In contrast, networks composed of long parallel filaments seem to be more dynamic and motile. It is not clear whether these differences in behaviour and physical properties stem from variations in the rates of filament turnover or flexibility, or perhaps from changes in cell adhesiveness, which might be affected by the Arp2/3 complex.

A possible caveat is that the observation of branched actin could be somewhat dependent on the technique used to prepare cells for electron microscopy. Resch *et al.*⁷ have observed that in cells prepared for frozen hydrated electron microscopy, the actin filaments in lamellipods seem much less branched than in cells prepared by critical-point drying (the technique used by Bear *et al.*). Resch *et al.* postulate that dehydration during the latter technique causes a collapse of actin networks, perhaps leading to filament-to-filament annealing. Moreover, after critical-point drying, even pure actin filaments (with no associated Arp2/3 complex) show 'branches'⁸. But branching has been seen 'live' under the light microscope in the presence of the Arp2/3 complex⁶, so branches do seem to be a real aspect of actin assembly induced by the Arp2/3 complex.

How might Mena/VASP promote more dynamic, faster-growing actin networks? Bear *et al.*¹ find that these proteins bind to the barbed (quickly growing) ends of filaments *in vitro*, and that barbed ends are

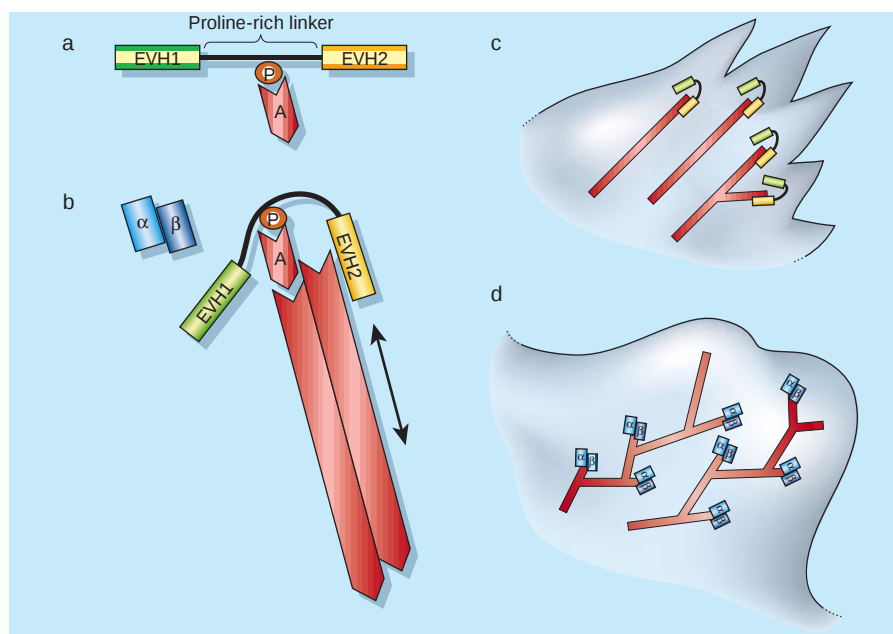


Figure 2 Model for movement. Bear *et al.*¹ find that Mena/VASP proteins promote the assembly of unbranched actin networks, leading to dynamic protrusions (lamellipods). **a**, A Mena/VASP protein, showing its various structural regions (domains). The EVH2 domain can interact with actin filaments. The proline-rich linker can interact with complexes of profilin (orange) and monomeric actin (red). **b**, Possible arrangement of Mena/VASP, bound to profilin-actin, near the barbed (quickly growing) end of an actin filament. The EVH2 domain might bind to the sides of actin filaments and slide along (double arrow) until it reaches the barbed end, where it would compete with capping protein (blue). **c**, Bear *et al.* show that Mena/VASP seems to block actin branching and capping in favour of rapid elongation, resulting in lamellipods that are highly dynamic, are shaped like sharks' teeth, and often lift off the surface, retract and change shape. **d**, In the absence of Mena/VASP, capping protein can bind to filament barbed ends, prohibiting further elongation. This leads to highly branched actin networks and dune-shaped lamellipods that are more conducive to cell movement than the sharks' teeth.

required for Mena/VASP to localize to the leading edges of lamellipods *in vivo*. This might explain how these proteins can reside in focal adhesions (complexes that link a cell's cytoskeletal filaments, such as actin, to its extracellular environment, and aid in movement) as well as at leading edges⁵ (Fig. 1). In any case, this and other evidence points to a model in which Mena/VASP is recruited to newly formed filaments at the leading edge, where it competes for filament barbed ends with 'capping' proteins (Fig. 2). If so, then filament dynamics might depend on the ratio of capping proteins to Mena/VASP. Capping proteins prevent filament elongation and so would promote shorter filaments. Mena/VASP would facilitate a longer growth phase, possibly through its ability to bind to profilin — an actin-monomer-binding protein that brings actin to uncapped barbed ends.

And how does Mena/VASP alter actin-filament branching? It is not yet clear why there are fewer branches in Mena/VASP-rich cells. There are many variables that affect the frequency, distribution and length of branches, and we are only just beginning to understand them. They include the possibility that the Arp2/3 complex competes with Mena/VASP for binding at or near barbed ends^{9,10}, and that filaments might in fact elongate more

rapidly (so increasing the spacing between branches) in the presence of Mena/VASP proteins than in their absence. Although there is only preliminary evidence for cellular factors that promote more rapid filament elongation¹¹, this has been a topic of speculation for many years. Future experiments should reveal the mechanism for the reduction in branching.

This clarification of the role of Mena/VASP proteins¹ has added more fine detail to our picture of how cells change shape and move. We now need to consider that the underlying architecture of actin-based protrusions is likely to be subtly different among superficially similar structures. In retrospect, it seems rather obvious that different cell types, facing diverse environmental challenges, will respond with slightly varied shape changes and protrusions. Looking ahead, however, it may seem daunting that cells appear to have many ways to take a walk — or even a tentative first step. ■

Laura M. Machesky is in the School of Biosciences, Division of Molecular Cell Biology, University of Birmingham, Birmingham B15 2TT, UK.
e-mail: l.m.machesky@bham.ac.uk

1. Bear, J. E. *et al.* *Cell* **109**, 509–521 (2002).
2. Renfranz, P. J. & Beckerle, M. C. *Curr. Opin. Cell Biol.* **14**, 88–103 (2002).
3. Reinhard, M., Jarchau, T. & Walter, U. *Trends Biochem. Sci.* **26**, 243–249 (2001).



100 YEARS AGO

Having regard to the wide reputation which the Malays have earned for themselves as a maritime people in Eastern seas, it is at first sight not a little remarkable that, so far as the Malay Peninsula is concerned, they have developed no really able type of sea-going boat. Three main factors have been at work influencing the development of boats, and tending to produce the characteristic shallow draft, lack of beam, and a consequent want of stability and weather lines. (1) The rivers are protected by very shallow bars of sand or mud, which make it impossible for a deep-bodied boat to obtain shelter within them. (2) The variable character of the light breezes prevailing in the Straits of Malacca. (3) The great strength of the tides. The lot of the sailing vessel is thus precarious; racing tides and baffling winds and calms make progress very slow. Hence propulsion by oars or paddles was the first necessity of the old-time Malay seaman in the Straits; sails were merely an occasional convenience.

From *Nature* 29 May 1902.

50 YEARS AGO

It has been established by earlier investigators that acetic acid has a destructive effect on the ascorbic acid in raw cabbage. This effect is somewhat surprising, since the lower the pH in the medium, the more stable is the ascorbic acid and, therefore, one would expect the acetic acid to have a preservative effect on the ascorbic acid in the cabbage. However, in experiments carried out in the early months of 1951, we found that, in many fruits and vegetables, the ascorbic acid is to a remarkable degree oxidized into dehydroascorbic acid if slices are sprinkled with 5 per cent acetic acid and allowed to stand for two hours ... in cabbage, cucumber, horse-radish, carrot, potato, lettuce, dill, leek, apple, pear and banana, 60–90 per cent of the ascorbic acid is oxidized. In parsley, spinach, cauliflower and tomato, the corresponding values are 20–50 per cent, and in orange and onion, only 0–10 per cent. We conceive the effect to be similar to that caused by mechanical damage to the cells. The acid penetrates into the cells and the hydrogen ions bring about a disturbance of the balance between the oxidizing and the reducing enzyme systems of the cell.

From *Nature* 31 May 1952.

4. Rottner, K., Behrendt, B., Small, J. V. & Wehland, J. *Nature Cell Biol.* **1**, 321–322 (1999).
5. Bear, J. E. *et al. Cell* **101**, 717–728 (2000).
6. Amann, K. J. & Pollard, T. D. *Proc. Natl Acad. Sci. USA* **98**, 15009–15013 (2001).
7. Resch, G. P., Goldie, K. N., Krebs, A., Hoenger, A. & Small, J. V. *J. Cell Sci.* **115**, 1877–1882 (2002).
8. Resch, G. P., Goldie, K. N., Hoenger, A. & Small, J. V. *J. Struct. Biol.* (in the press).
9. Ichetovkin, I., Grant, W. & Condeelis, J. *Curr. Biol.* **12**, 79–84 (2002).
10. Pantaloni, D., Boujemaa, R., Didry, D., Gounon, P. & Carlier, M. F. *Nature Cell Biol.* **2**, 385–391 (2000).
11. Huang, M. *et al. Curr. Biol.* **9**, 979–982 (1999).

Earth science

The slippery slope

George Zandt

Seismic readings suggest that a zone of weak, slippery rocks lies beneath the Pacific northwest coast of the United States. These frail layers might be limiting the violence of earthquakes.

Most of our planet's largest earthquakes and active volcanoes are associated with subduction zones, where one tectonic plate dives under another. As the downgoing plate experiences higher pressures and temperatures, water trapped in its crust and upper mantle is gradually forced out and into the mantle of the overlying plate. At depths of about 100 km, the addition of water lowers the melting temperature of the mantle rocks and initiates partial melting to magma, which rises and erupts in the so-called arc volcanoes that form the 'Ring of Fire' along the Pacific Rim.

Reporting on page 536 of this issue, Bostock and colleagues¹ reveal seismic images of the Cascadia subduction zone, which runs along the Pacific northwest coast of the United States. Using these images, they investigate what happens when water moves into the shallow mantle in front of the volcanic arc (Fig. 1). They show that the forearc mantle, located 40–80 km deep between the coast and the volcanic arc, is largely 'serpentinized'. This means that it has undergone a process that changes the mineralogy of the mantle rocks and their physical properties — a transformation that could have important bearing on the occurrence of earthquakes around subduction zones.

Serpentinization, which can occur when water moves into cool mantle rocks, chemically changes the strong, dry rocks into weak, hydrous rocks — or serpentinite — composed of serpentine minerals, talc and brucite². Chrysotile, the chief source of asbestos, is a serpentine mineral with a fibrous structure. Antigorite, in contrast, is a serpentine mineral that has a plate-like structure and is stable at the temperatures found in the forearc mantle. It is the fibrous or plate-like structures of these minerals that give serpentinites a slippery feel. And although the fibrous nature of asbestos makes it a dangerous health hazard in the environment, slippery serpentinite in subduction zones might be important in limiting the depth at which earthquakes are generated.

Bostock *et al.*¹ located the serpentinite deep under the Cascadia subduction zone when they found an 'inverted Moho' beneath the forearc region. A 'Moho' is the boundary at which the speed of seismic waves rises to a value typical for waves travelling through mantle rocks (around 8 km s^{-1}), and is named after the Croatian Andrija Mohorovičić, who discovered the phenomenon in 1910. Usually, a Moho is associated with a change in chemical composition from crustal to mantle rocks that occurs, on average, at depths of around 40 km under the continents. But Bostock *et al.*, remarkably, have seen an inverted Moho. Under the high Cascade mountains, at a depth of 36 km, the Moho is a normal inter-

face with seismic wave speed increasing with depth. As the authors traced the Moho laterally towards the Pacific Ocean, they saw the interface flip over so that seismic wave speed decreases with depth (Fig. 1).

Bostock *et al.* suggest that this unusual Moho inversion is caused by a major reduction in the wave speed through the mantle — to lower the speed to this extent would require a level of serpentinization of about 50–60%. Much of the water needed for this serpentinization might come from another chemical reaction associated with subduction. As the oceanic crust of the downgoing plate comes under increasing pressure, it transforms into eclogite, a dry high-density rock with seismic wave speeds similar to those in mantle rocks. The seismic images taken by Bostock *et al.* show that the oceanic Moho diminishes, as the crust transforms into eclogite, at a depth of about 45 km, which is consistent with the idea that dehydration of the oceanic crust hydrates the overlying forearc mantle.

Serpentinite is stable to temperatures of 600–700 °C, but at higher temperatures it breaks down into other minerals and water. Bostock *et al.*¹ have devised a thermal model for the Cascadia subduction zone and show that the forearc mantle is sufficiently cold for serpentinite to be stable. But higher mantle temperatures are predicted to exist beneath the volcanic arc, consistent with the reversal of the inverted Moho seen in the seismic image and with the onset of volcanism. If the

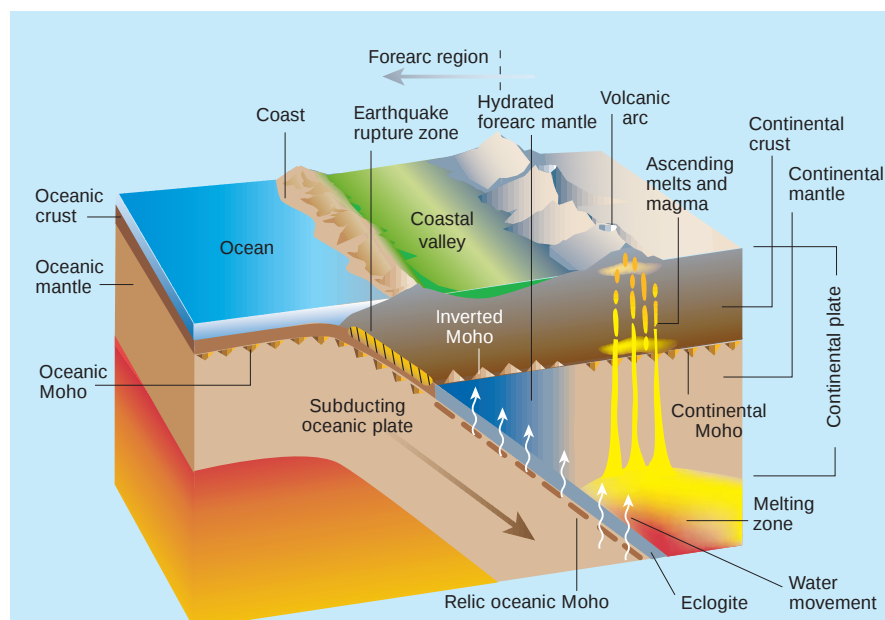


Figure 1 Cross-section through the Cascadia subduction zone on the northwest coast of the United States. Water released from the subducting oceanic crust rises up and hydrates the forearc mantle; water released deeper down causes partial melting of rock to magma that erupts in the arc volcanoes. In the water-altered — or 'serpentinized' — forearc rocks, seismic waves travel more slowly and the usual gradient of wave speed is reversed, producing an effect known as an 'inverted Moho', which has been detected by Bostock *et al.*¹. The slippery nature of the serpentinized forearc mantle limits the seismogenic depth of coupling between the two plates, and so might also limit the impact of earthquakes in the region.

oceanic crust is dry (having given up its water to serpentinization) by the time it reaches the depths at which the mantle melts, then from where does the water to initiate mantle melting come? It could be coming from the mantle of the downgoing slab, or perhaps the lower portion of the serpentinized mantle is being dragged down and is releasing its waters at greater depths³. However, the low density and slippery nature of the serpentine may work against this latter mechanism.

Serpentinized forearc mantle has been found in at least two other subduction zones: Japan and the Central Andes of South America. In the Central Andes, seismic studies have identified a 'blurred' Moho, indicative of hydration⁴, as well as uncovering evidence that dehydration reactions affect earthquakes in the subducting plate^{5,6}. If a fully serpentinized forearc mantle contains about 6% by weight of water⁷, then the total volume of water held there is about 540 km³ per linear km of subduction zone; and if all 37,000 km of the subduction zones on Earth are 50% hydrated, then the total volume of water is 10⁷ km³ — or about one-third of the present volume of water in the Antarctic ice cap. So, if shown to be present globally, the serpentinized forearc mantle could prove to be a substantial water reservoir in the global hydrologic cycle.

One of the most important implications of serpentinization of the forearc mantle is

the possible effect on the termination of large earthquake ruptures². Seismologists have found that these ruptures often terminate near the depth of the continental Moho⁸. The slippery nature of serpentine may provide a ready explanation for this observation. If these earthquake ruptures occurred deeper in the subduction zone, below the serpentinized layer, they would produce strong shaking over a greater area, threatening a larger proportion of the coastal population centres. Thus, the hydrated forearc mantle may be pivotal in protecting the coastal cities of the US Pacific northwest, as well as other major cities around the Pacific Rim, from greater earthquake danger. ■

George Zandt is in the Department of Geosciences, University of Arizona, Tucson, Arizona 85721-0077, USA.

e-mail: zandt@geo.arizona.edu

1. Bostock, M. G., Hyndman, R. D., Rondenay, S. & Peacock, S. M. *Nature* **417**, 536–538 (2002).
2. Peacock, S. M. & Hyndman, R. D. *Geophys. Res. Lett.* **26**, 2517–2520 (1999).
3. Bose, K. & Ganguly, J. *Earth Planet. Sci. Lett.* **136**, 109–121 (1995).
4. Giese, P., Scheuber, E., Schilling, F., Schmitz, M. & Wigger, P. *J. S. Am. Earth Sci.* **12**, 201–220 (1999).
5. Yuan, X. et al. *Nature* **408**, 958–961 (2000).
6. Bock, G., Schurr, B. & Asch, G. *Geophys. Res. Lett.* **27**, 3929–3932 (2000).
7. Schmidt, M. W. & Poli, S. *Earth Planet. Sci. Lett.* **163**, 361–379 (1998).
8. Ruff, L. J. & Tichelaar, B. W. in *Subduction: Top to Bottom* (eds Bebout, G., Scholl, D. W., Kirby, S. H. & Platt, J. P.) 105–111 (AGU, Washington DC, 1996).

Miocene in a novel way. They tabulated hypsodonty for 209 European fossil localities by assigning each species a value corresponding to low-, medium- or high-crowned teeth, and then examined the numerical trend in those values for common animals (found in 25% or more of the localities) and rare animals (found in less than 25% of localities). Importantly, they did not compare relative abundances within localities, so that their definition of common does not imply that these animals were the most abundant ones in any given locality, but that they occurred at many localities across the region in each time interval. Rarity was assessed for each interval, so groups that were rare during one time were not necessarily rare at another.

Jernvall and Fortelius examined changes in hypsodonty for species and genera over time, and found similar patterns at both taxonomic levels. In rare genera, tooth crown height stayed relatively constant. In the more geographically widespread ones, however, average hypsodonty began to increase significantly about 11 million years ago. To ensure that this pattern was not driven by particular lineages, the authors selectively removed potentially problematic groups (horses, primates and proboscideans) from the analysis: horses because in the textbook version of events they are the group that were thought to drive the trend of increasing hypsodonty; primates because they have low-crowned teeth and are relatively rare; and proboscideans because they are large and their fossils are usually more likely to be collected than those of small taxa. The pattern was still present in the remaining data.

Brachydont species are often smaller than hypsodont ones, so Jernvall and Fortelius examined the degree of hypsodonty in different crown morphology types. Crown types are characterized by functional properties of the teeth, such as the number, orientation and shape of the cusps³, and should be insensitive to size differences among species. The trend of increasing hypsodonty is also found only among geographically common crown types, supporting the view that this trend is not a taxonomic artefact. The proportion of localities that include hypsodont genera and crown types increased significantly after 11 million years ago, whereas the proportion of localities with low-crowned genera and crown types did not. These results imply that drying climatic conditions in Europe, usually associated with the uplift of the Tibetan plateau⁶, allowed ungulates — hoofed mammals — with higher tooth crowns to expand their range as more open arid grasslands spread in Europe. Work is needed to see whether the trend in the North American fossil record, which has better temporal resolution, is also driven by common taxa. Examination of the relative abundances of taxa within localities

Evolutionary biology

Crowning glories

Jessica M. Theodor

The evolutionary trend towards high-crowned teeth in European herbivores during the Miocene was, it seems, driven by the geographically common groups, and not rare ones. That conclusion has broader implications.

The evolution of high-crowned teeth is one of the classic stories of evolutionary biology, but the object lesson of the story might be different from that found in the textbooks. In studying past environments, the rare and more specialized animals often get the most attention. In their paper on page 538 of this issue, however, Jernvall and Fortelius¹ show that a few, common, regionally widespread species can drive major trends, yielding more information than is usually recognized.

Twenty-four million years ago, at the beginning of the Miocene epoch, most herbivorous mammals had low-crowned (brachydont) cheek teeth and a few rare species had high-crowned (hypsodont) teeth. Hypsodont teeth are an adaptation to increased abrasive wear. Although high-crowned teeth are no more mechanically resistant to abrasion than low-crowned ones, the additional height of the tooth

crown means that there is more tooth to wear away over the animal's lifetime. However, these teeth are costly in terms of both the materials needed to produce them and the shape changes that must evolve in the jaws to accommodate them.

Hypsodont teeth have evolved in many herbivorous lineages in response to increased abrasion, whether by the silica secreted by grasses² or by grit in the environment stuck to plant foods^{3,4}. By the end of the Miocene, 5 million years ago, the situation had reversed compared with the beginning of the epoch, with hypsodont forms present at most localities. This change is attributed to increased aridity and the spread of grasslands during this time, which would increase abrasive wear by the spread of more open habitats with higher grit levels, and also by the spread of silica-bearing grasses.

Jernvall and Fortelius¹ examined changes in tooth crown height during the

in both continents should show whether this pattern is unique to Europe or is more widespread.

Jernvall and Fortelius point out that their results are consistent with the life-history differences we see in modern ungulates. Hypsodont species are somewhat larger, and are more likely to form large herds⁷ and disperse more widely across landscapes, and the increase in hypsodonty among common taxa about 11 million years ago might herald the evolution of these life-history traits. But showing that these traits are associated with a trend of increasing hypsodonty in common herbivores, and are not found in the rarer ones, will require more work on the Miocene fossil record.

Aspects of Jernvall and Fortelius's approach will influence future work on evolutionary trends. The use of a new metric — average hypsodonty — could prove useful as a palaeoenvironmental indicator, and provides a quantitative measure of the temporal and spatial variation in crown heights of teeth from different sites. The examination of common and rare groups has allowed the

finer resolution of a major trend and might lead to a better understanding of the evolution of grassland ecosystems. It also suggests that common groups yield more information about palaeoecology than has previously been recognized. The use of large data sets to examine evolutionary trends has increased markedly in recent years, and further insights into other trends, such as changes in body size, might result from examining common and rare taxa separately.

Jessica M. Theodor is in the Department of Geology, Illinois State Museum, 1011 East Ash Street, Springfield, Illinois 62703, USA.
e-mail: jtheodor@museum.state.il.us

1. Jernvall, J. & Fortelius, M. *Nature* **417**, 538–540 (2002).
2. Jacobs, B. F., Kingston, J. D. & Jacobs, L. L. *Ann. Missouri Bot. Gdn* **86**, 590–643 (1999).
3. Janis, C. M. & Fortelius, M. *Biol. Rev. Camb. Phil. Soc.* **63**, 197–230 (1988).
4. Janis, C. M. in *Teeth Revisited: Proceedings of the VIIth International Symposium on Dental Morphology* (eds Russell, D. E., Santoro, J.-P. & Sigogneau-Russell, D.) 367–387 (Musée d'Histoire Naturelle, Paris, 1988).
5. Jernvall, J., Hunter, J. P. & Fortelius, M. *Science* **274**, 1489–1492 (1996).
6. Ruddiman, W. F. (ed.) *Tectonic Uplift and Climate Change* (Plenum, New York, 1997).
7. Jarman, P. J. *Behavior* **48**, 215–267 (1974).

Nuclear physics

Weighing up nuclear masses

Sven Åberg

It is difficult to match theoretically calculated masses of atomic nuclei to their experimentally measured values. It seems that chaotic motion inside the nucleus may be the reason for this discrepancy.

The atomic nucleus is central to many physical and astrophysical processes, including the nuclear processes that power the Sun and synthesize the elements. Which atoms exist depends on the stability of different types of nuclei, and the stability of an individual nucleus depends on its mass. In theoretical nuclear physics, it has long been a challenge to calculate the masses of known (and as yet unknown) nuclei — but experimentally measured and theoretically calculated masses do not fully agree. Writing in *Physical Review Letters*, Bohigas and

Leboeuf¹ suggest that chaotic motion in the nucleus could influence nuclear mass, and could ultimately limit how accurately nuclear masses can be calculated theoretically.

The atomic nucleus consists of neutrons and protons held together by the strong nuclear force. Mathematically, this complicated 'many-body' problem can be described fairly accurately by assuming that the interaction of a given particle inside the nucleus with all other particles can be approximated by an averaged energy term, the 'mean-field potential'. Einstein equated energy and mass

in his famous equation, $E = mc^2$, so information about the mean-field potential can be obtained from studies of nuclear masses.

When nuclear mass is studied as a function of the number of protons or neutrons inside the nucleus, bunching and gaps in the quantum spectrum of the mean-field potential are revealed. This variation, called 'shell structure', is a general phenomenon in quantum systems with a small number of particles — such as electrons in an atom or a quantum dot, or clusters of atoms. The gaps in the quantum spectrum arise as particles fill up closed shells of quantum-mechanically allowed orbits. In the spectrum of nuclear masses, local minima suggest energetically favourable configurations of neutrons (and protons), in which the number of neutrons (or protons) is sufficient to complete a shell (Fig. 1).

The shell energy is a quantum-mechanical quantity, but in principle it can be expressed in classical terms, by summing up all possible periodic orbits for a particle as it moves in the classical mean-field potential^{2,3}. Bohigas and Leboeuf¹ consider the dynamics of the atomic nucleus as being made up of layers of regular motion, overlaid with chaotic motion. The shell energy then depends on contributions from regular orbits and from chaotic orbits. Provided that the two contributions are independent, the shell energy can be split into the two parts, each expressed in terms of asymptotic series of classical periodic orbits, with fluctuations calculated analytically^{1,3}. The contribution from chaotic orbits can then be compared with the difference between experimentally measured and calculated shell energy, δE (Fig. 1), and the calculated shell energy can be compared with the contribution from regular orbits.

Although the fluctuation of calculated shell energies for all nuclei is about 3.5 MeV, the fluctuation in δE is considerably smaller, averaging about 0.7 MeV. The discrepancy between measured and calculated masses decreases with increasing neutron number (Fig. 1). In their analysis, in terms of regular and chaotic orbits, Bohigas and Leboeuf reproduce this behaviour, supporting the interpretation of the deviation between

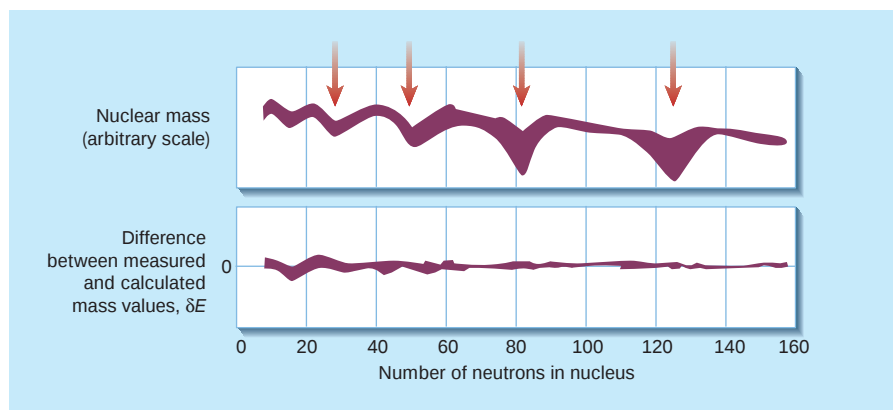


Figure 1 Nuclear mass varies with the number of neutrons in the nucleus, showing a 'shell structure' (upper panel). The minima at neutron numbers 28, 50, 82 and 126 correspond to especially stable nuclei, in which the number of neutrons is sufficient to complete a shell (a band of masses is shown because, for a given number of neutrons, several nuclei with different numbers of protons can exist). Theoretical models cannot quite describe this variation, and a discrepancy is always seen between calculated and measured mass values (lower panel). Bohigas and Leboeuf¹ suggest that this can be explained by taking into account chaotic motion inside the nucleus. (Graphic derived from ref. 5.)

measured and calculated energies as resulting from chaotic motion. It seems that chaotic motion has to be considered if theoretical descriptions of nuclear masses are to be improved.

A plausible explanation for the appearance of chaotic motion is that there are many-body effects that are not included in the mean-field potential. Such effects give rise to very complicated spectra when nuclei are excited to neutron or proton resonances. The effort to explain these complicated spectra led Wigner in the 1950s to introduce a theory — random matrix theory⁴ — that is now a key concept of quantum chaos. According to this theory, properties related to individual energies can be calculated only in a statistical sense. If this is the situation for nuclear masses as well, it will certainly be extremely hard to improve the theoretical description of nuclear masses. Even in the best theoretical models^{5–7}, it is exceedingly difficult to keep the model uncertainties lower than about 0.7 MeV, the level of the expected shell-energy contribution from chaotic motion.

In their calculations, Bohigas and Leboeuf¹ have assumed that there are no correlations between chaotic and regular layers of motion. But the existence of non-zero correlations between the calculated shell energy and the difference in measured and calculated values (a correlation of about –16% is

obtained from ref. 5) — as well as some uncertainties in estimating fluctuations of the chaotic contribution to shell energy — suggests that there is room for improvement before the chaotic limit is reached.

Indeed, much is being done in theoretical nuclear-physics research to try to improve the description of nuclear masses⁷, some going beyond the mean-field description^{8–10}. Such efforts could allow us to extrapolate more effectively into nuclear-mass regions that have yet to be explored — in particular, regions where the astrophysical creation of elements takes place. ■

Sven Åberg is in the Department of Mathematical Physics, Lund Institute of Technology, PO Box 118, S-221 00 Lund, Sweden.

e-mail: sven.aberg@matfys.lth.se

1. Bohigas, O. & Leboeuf, P. *Phys. Rev. Lett.* **88**, 092502 (2002).
2. Gutzwiller, M. C. *Chaos in Classical and Quantum Mechanics* (Springer, New York, 1990).
3. Leboeuf, P. & Monasteria, A. G. *Ann. Phys.* **297**, 127–156 (2002).
4. Wigner, E. P. in *Statistical Theories of Spectra: Fluctuations* (ed. Porter, C. E.) 88–140, 145–161, 176–180, 188–211, 223–227, 446–461 (Academic, New York, 1965).
5. Möller, P., Nix, J. R., Myers, W. D. & Swiatecki, W. J. *Atom. Data Nucl. Data Tables* **59**, 185–382 (1995).
6. Goriely, S., Tondeur, F. & Pearson, J. M. *Atom. Data Nucl. Data Tables* **77**, 311–381 (2001).
7. Samyn, M., Goriely, S., Heenen, P.-H., Pearson, J. M. & Tondeur, F. *Nucl. Phys. A* **700**, 142–156 (2002).
8. Otsuka, T. *Hyperfine Inter.* **132**, 409–416 (2001).
9. Valor, A., Heenen, P.-H. & Bonche, P. *Nucl. Phys. A* **671**, 145–164 (2000).
10. Rodriguez-Guzman, R., Egido, J. L. & Robledo, L. M. *Phys. Lett. B* **474**, 15–20 (2000).

Ion channels

An open and shut case

Maria Schumacher and John P. Adelman

Most ion channels open and close — they are ‘gated’ — in response to cues in their environment. A crystal structure of a Ca^{2+} -gated K^+ -ion channel provides insight into how gating works.

Nervous systems are built for speed, using electrical signals that are propagated by the activities of ion channels. Two responsibilities are incumbent on these membrane-spanning proteins. First, they must provide a selective conduction pathway, the pore — a hospitable environment through which appropriate ions can be escorted across a cell’s outer membrane. Second, most ion channels must undergo rapid conformational changes in response to metabolic cues, opening or closing the pore. These ‘gating’ cues can be changes in voltage across the membrane, or molecules (ligands), or both. Just how ion channels conduct ions, how ligands bind to channel proteins, and how gating cues are transduced into conformational rearrangements are the biggest questions facing this field. Some answers are provided by two papers from Roderick MacKinnon’s laboratory, on pages 515 and 523 of this issue^{1,2}. In the first paper the authors describe the crystal structure of

a Ca^{2+} -gated K^+ channel, revealing details of the Ca^{2+} -occupied ligand-binding region and the open pore¹. In the second they provide the first comparison of the structures of open and closed K^+ -channel pores, and offer a general model for gating².

MacKinnon’s group has presented a series of stunning papers over the past four years, taking advantage of the ability to obtain large quantities of purified ion-channel proteins from bacteria. Crystal structures of closed KcsA channels (K^+ channels from *Streptomyces lividans*) containing resident K^+ ions revealed the essential features of K^+ conduction^{3–5}. These channels consist of four protein subunits, and each subunit contributes one membrane-spanning segment (a helical structure) to the lining of the pore. At the interface of the channel with the cytoplasm inside the cell, the helices come together to form a crossed bundle rather like an inverted teepee, which makes up the physical gate of the channel. Farther into the

membrane is an inner cavity that perfectly accommodates a hydrated K^+ ion. This cavity is found just beneath the selectivity filter, which discriminates between K^+ and other ions. Four rings of backbone carbonyls form the filter, and during dynamic conduction they are probably occupied alternately by water and K^+ ions. In short, every aspect of KcsA seems to have evolved for K^+ -specific conduction.

What about channel gating? Elegant database analysis⁶ suggested that an intracellular structural domain shared by many bacterial K^+ channels and other transport systems, and also found in mammalian channels that are gated by the combination of Ca^{2+} and voltage, is important for gating. Previous work⁶ also hinted that this domain, called RCK for ‘regulates conduction of K^+ ions’, might be a ligand-binding motif.

The latest work from MacKinnon’s group brings our knowledge of gating and conduction together. In the first paper¹ they present the structure of a Ca^{2+} -activated K^+ channel (MthK, from *Methanobacterium thermoautotrophicum*), with its pore and attached RCK domains. The structure shows that functional K^+ channels consist of four pore-forming subunits, as well as four RCK dimers (that is, eight RCK domains) that form a ‘gating ring’ just inside the cytoplasm (Fig. 1, overleaf). Two important interfaces between the RCK domains keep the ring in place. One interface is fixed and probably occurs in all types of RCK domain. The other is flexible, made up in part by subdomains that protrude from each RCK domain and form a ligand-binding pocket between adjacent domains; in the crystal structure, two Ca^{2+} ions are clearly positioned at the base of this cleft between two RCK domains.

In the same paper, the authors compare the Ca^{2+} -bound RCK domain from MthK with the unbound RCK domain from an *Escherichia coli* K^+ channel, providing insights into how Ca^{2+} binding might exert mechanical force to open the attached pore-lining helices. The fixed interfaces in the two structures are similar, indicating that Ca^{2+} does not alter their relative orientations — they are indeed fixed. But the flexible interfaces are remarkably different. It seems likely that Ca^{2+} binding reshapes these clefts, and that this compels a tilt of the rigid interfaces, expanding the diameter of the gating ring and exerting a force on the attached pore-lining helices that opens the pore (Fig. 1). In this way, MthK uses the chemical energy of Ca^{2+} binding to perform the mechanical work of gating.

A special consequence of the structure of MthK is that it captures the pore in an open state, and this has allowed the authors to compare it with the previously determined closed-pore structure of KcsA. As they describe in the second new paper², the contrast is remarkable. The outer region of the

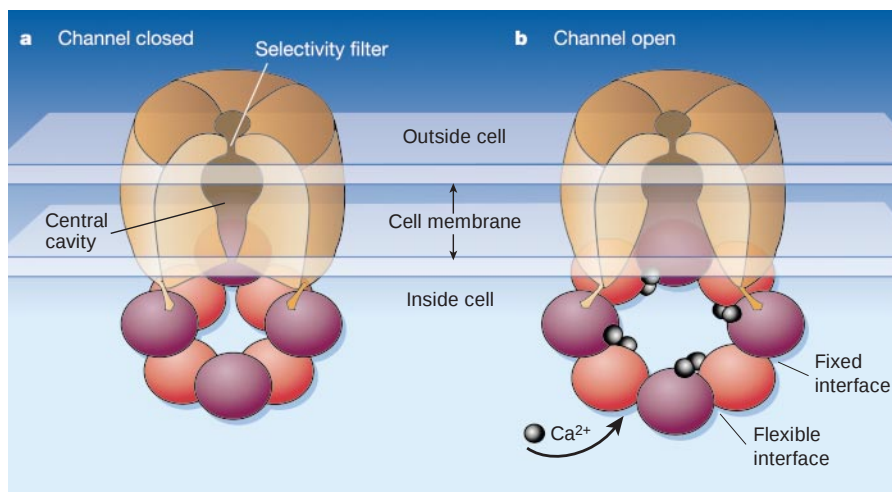


Figure 1 Schematic model for the gating and opening of a bacterial Ca^{2+} -gated K^+ channel, based on the new papers by MacKinnon and colleagues^{1,2}. **a**, The closed conformation; **b**, the open conformation. Three of the four subunits of a K^+ channel are shown in brown. The purple and red circles represent the eight RCK domains, which, after binding Ca^{2+} , are thought to reorientate with respect to each other, causing changes in the pore in the centre of the channel.

pore, including the selectivity filter, is very similar in the two structures. But in the closed KcsA pore the helices of the 'teepee' are nearly straight, whereas they are bent and splayed open in the open MthK pore — and the separate identity of the central cavity is essentially lost.

The hinge of the helices occurs at a particularly provocative position surprisingly deep within the pore, at a glycine amino acid close to the selectivity filter. This makes perfect sense, given the predicted consequences of Ca^{2+} binding: the reorganization of the gating ring exerts an outward, radial force that focuses at the glycine, taking advantage of the flexibility endowed by this amino acid and causing the bend. Sequence comparisons² of the pore-lining inner helices from a wide array of K^+ -selective channels reveal glycine at an analogous position and an alanine strategically positioned five amino acids carboxy-terminal to the glycine, suggesting that the gating mechanism is conserved across many different channels. The combination of a wide inner cavity and a narrow selectivity filter effectively focuses the membrane to the length of the selectivity filter, some 12 Å, probably explaining the high conduction rates of K^+ channels.

The potential diversity and intriguing similarities of gating mechanisms can be highlighted by comparing MthK with the structure⁷ of the ligand-binding domain from a mammalian Ca^{2+} -activated K^+ channel (the SK channel) without the associated pore. In the SK channel, a ligand-binding domain that is unrelated to the RCK domain yet is similarly positioned just beneath the membrane, retains bound calmodulin — a protein that serves as the Ca^{2+} -binding unit. A conformational change after Ca^{2+} binding, involving dimerization of the ligand-binding domain, apparently exerts a

force on the attached inner helices and opens the pore⁸.

So both channels apparently use a chemomechanical mechanism⁷ to couple Ca^{2+} binding to conformational rearrangements that open the pore. Future studies

should reveal whether SK channels control their pores according to the model proposed for MthK; this seems plausible, as the mammalian SK channels retain the strategic glycine-alanine motif. Interestingly, in both of these structures^{1,7}, the 17-amino-acid stretch that links the pore domain with the Ca^{2+} -bound ligand-binding domain was not resolved. This stretch may be the focus for the mechanical force produced by Ca^{2+} binding. Structures of ligand-gated channels that resolve this important linker may provide yet more surprising mechanistic insights.

Maria Schumacher is in the Department of Biochemistry and Molecular Biology, and John P. Adelman is at the Vollum Institute, Oregon Health and Sciences University, Portland, Oregon 97201, USA.

e-mail: adelman@ohsu.edu

1. Jiang, Y. *et al.* *Nature* **417**, 515–522 (2002).
2. Jiang, Y. *et al.* *Nature* **417**, 523–526 (2002).
3. Doyle, D. A. *et al.* *Science* **280**, 69–77 (1998).
4. Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. *Nature* **414**, 43–48 (2001).
5. Morais-Cabral, J. H., Zhou, Y. & MacKinnon, R. *Nature* **414**, 37–42 (2001).
6. Jiang, Y., Pico, A., Cadene, M., Chait, B. T. & MacKinnon, R. *Neuron* **29**, 593–601 (2001).
7. Schumacher, M. A., Rivard, A. F., Bachinger, H. P. & Adelman, J. P. *Nature* **410**, 1120–1124 (2001).
8. Xia, X.-M. *et al.* *Nature* **395**, 503–507 (1998).

Neurobiology

Model hearing

George D. Pollak

For the past 50 years a particular model of how animals locate the source of sounds has driven much of the research on auditory systems. It now seems, however, that this model might not apply to mammals.

Our ability to pinpoint the source of the sounds we hear is quite remarkable, but how do we do it? In general terms, the answer has been known since the time of Lord Rayleigh¹: the localization of low-frequency sounds relies on a neural computation based on interaural time disparities (ITDs) — differences in the time it takes for a sound to reach each ear. Incredibly, we can detect ITDs of only a few microseconds, allowing us to distinguish between sounds separated by only a few degrees in space².

In 1948 Lloyd Jeffress³ proposed a model that explains how the auditory nervous system processes ITDs and how it represents the particular ITD that is received at the ears. The model is charming in its simplicity and elegance, and has driven much of the research on vertebrate auditory systems for the past 50 years. But Brand and colleagues⁴ (page 543 of this issue) now provide evidence for a very different model. They show that the timing of neuronal inhibition — a feature not incorporated in the Jeffress model — shapes ITD processing

and representation in previously unsuspected ways.

The Jeffress model for the processing and representation of ITDs has enormous appeal because it incorporates neural processes and structure to account for a behaviour — sound localization. Jeffress envisaged arrays of binaurally innervated neurons, with the neurons in each array presumably all tuned to the same frequency. Each neuron receives an excitatory input from one ear via axons (projections) of a particular length, and an input from the other ear via axons of a slightly different length. The difference in axon lengths creates a delay line, and thus a difference in the times at which action potentials from the two ears arrive at the binaural neuron. This time difference compensates for the ITD between the ears.

The central idea of the Jeffress model is that each neuron fires maximally only when action potentials from the two ears are coincident, arriving at the binaural neuron at the same time. As the delay lines are different for

Daedalus

Spatial audio

In electron spin resonance (ESR), a chemical sample is placed in a strong magnetic field and irradiated with microwaves. Unpaired electrons in the sample reveal themselves by resonating. Daedalus has been musing on interstellar space, and its content of hydrogen atoms. Each has an unpaired electron (indeed, it is the only one it has). In the low magnetic field of space, each electron should resonate feebly at some audio frequency. Yet the total signal might be quite strong: there are cubic light-years of this specimen. Indeed, the distribution of monatomic hydrogen, and the range and intensity of the interstellar magnetic field itself, are all hot astronomical topics. Daedalus reckons that audio frequencies from space are well worth looking for.

An audio signal would have a very long wavelength — 200 km or thereabouts. A directional parabolic aerial, which must be many wavelengths across, could never be made big enough. Even a conventionally resonant quarter-wave aerial of 50 km would be hard to build. But Daedalus recalls that power-lines and telegraph cables, thousands of kilometres long, already span the globe. Furthermore they are liable to dangerous surges when solar magnetic effects induce big voltages in them. This astrophysical phenomenon suggests to him that a careful search should be made on these conductors, looking for another astrophysical effect: small but detectable audio frequencies from space.

Of course 50 Hz and 60 Hz, the main human power frequencies, will contribute hugely, and will have to be well filtered out. These, however, will have their own human rhythm, caused by the known changing load. The audio spectrum to be studied is wide, too. Furthermore, the system as a whole will be steadily scanned in longitude by the rotation of the Earth, and in latitude by choosing the right conductors to listen to. Circumpolar ones, for example, should discriminate rather well against equatorial signals or ones from the wrong hemisphere.

Here, says Daedalus, is a new way of testing the theory of 'steady-state continuous creation'. It holds that monatomic hydrogen is appearing steadily throughout space, at just the rate needed to compensate for spatial expansion. The Universe has no beginning and no end. If the Earth is indeed receiving a steady ESR hydrogen signal from every point in space, the implications would be cosmologically profound indeed.

David Jones

each neuron, a sound from a particular location, which generates a particular ITD, will produce coincident inputs in only one neuron (or a subset of neurons); all other neurons will receive non-coincident inputs and fire weakly. Changing the location of the sound, and hence the ITD, results in coincidence at a different neuron, and thus at a different place in the array. In this way, the ITD received is encoded by the place in the array at which neurons fire maximally. This arrangement is reiterated for each frequency.

So, for example, a sound emanating from directly ahead will generate an ITD of 0 μ s, and coincidence will occur only at those auditory neurons with equal axon lengths. Neurons in each frequency array that are 'tuned' to 0 μ s will fire maximally, and all other neurons will fire weakly. In this way, a complex sound originating directly in front of the animal will evoke maximal activity at one place across frequency arrays.

These features appear in the first brain region innervated by the two ears: the medial superior olive (MSO) in mammals and the nucleus laminaris in reptiles and birds. As early as 1953, Stotler⁵ reported a striking binaural innervation of MSO neurons. Subsequent neurophysiological studies all showed that these neurons respond to the coincidence of excitatory inputs and are exquisitely sensitive to ITDs in the microsecond range — properties consistent with the Jeffress model⁶. Even more compelling are studies of the nucleus laminaris in birds. There, the Jeffress-type delay lines have been demonstrated anatomically⁷, and the place coding of ITDs, resulting from coincidence of excitatory inputs, has been shown with physiological recordings^{8–10}.

But Brand *et al.*⁴ now provide evidence that the MSO in mammals might not work as previously supposed. The authors recorded the electrical activity of MSO neurons in gerbils, animals that are known to localize low-frequency sounds by means of ITDs, and confirmed that each neuron fires maximally at a particular ITD. Yet — and here's the rub — the peak firing occurred at long ITDs that gerbils almost never experience, because their headwidths are too small to generate them. So it is difficult to imagine that ITDs are indeed represented by the place code envisaged by Jeffress.

Brand *et al.* also report that the ITDs that produce maximal firing are closely correlated with the frequency to which each neuron is tuned. Neurons that fire maximally at relatively small ITDs are almost always tuned to higher-frequency sounds, whereas neurons that fire maximally at progressively longer ITDs 'prefer' progressively lower frequencies. So the peak firings generated by a particular ITD are not represented equally in each frequency array of the MSO, as predicted in a Jeffress-like arrangement. Instead, each frequency array responds to a small range of

ITDs over which the neuronal firing rate changes markedly.

This leads Brand *et al.* to suggest that, in mammals, the location of a sound is encoded not by the place of maximal firing but rather by the activity pattern across the entire population of auditory neurons, with these neurons changing their firing rates as ITDs change. In their model, high frequencies, which generate the steepest changes in firing rate with ITD, are scaled to peak at small ITDs so that the steep rate changes occur within the biologically relevant ITD range. Conversely, low frequencies, which generate the shallowest rate changes with ITD, are scaled to peak at longer ITDs so that the largest changes in firing also occur within the biologically relevant ITD range.

But the most surprising result occurred after Brand *et al.* blocked inhibitory inputs to MSO neurons. They found that all neurons tested, regardless of their usual behaviour, now fired maximally at ITDs of and around 0 μ s. Although the authors blocked inhibition in only a few MSO neurons, their results suggest that delay lines of excitatory inputs are not arranged to produce a range of ITD sensitivities, as Jeffress proposed. Rather, the implication is that, regardless of their arrangement, all neurons fire maximally at ITDs of 0 μ s. Inhibition, or more specifically its timing relative to excitation, then sculpts a variety of ITD sensitivities out of the common 0- μ s sensitivity produced by the excitatory inputs.

This work raises several questions. Why do mammals and birds have different mechanisms for localizing sounds? And what structural features underlie the inhibitory delay lines suggested by Brand *et al.*? Finally, gerbils are small mammals with small headwidths, and their peak neuronal firing occurred at ITDs that their headwidths could not generate. But what about larger mammals, whose headwidths generate much longer ITDs — does inhibition have a role to play here, too? All in all, the study by Brand *et al.* will no doubt generate considerable discussion about mechanisms that many had thought were already solved. ■

George D. Pollak is in the Department of Neurobiology, University of Texas, Austin, Texas 78712, USA.

e-mail: gpollak@mail.utexas.edu

1. Rayleigh, Lord. *Phil. Mag.* **13**, 214–232 (1907).
2. Klumpp, R. *et al.* *J. Acoust. Soc. Am.* **28**, 215–232 (1956).
3. Jeffress, L. A. *J. Comp. Physiol. Psychol.* **41**, 35–39 (1948).
4. Brand, A., Behrend, O., Marquardt, T., McAlpine, D. & Grothe, B. *Nature* **417**, 543–547 (2002).
5. Stotler, W. J. *Comp. Neurol.* **98**, 267–285 (1953).
6. Joris, P. X., Smith, P. H. & Yin, T. C. *Neuron* **21**, 1235–1238 (1998).
7. Parks, T. N. & Rubel, E. W. *J. Comp. Neurol.* **164**, 435–448 (1975).
8. Carr, C. E. & Konishi, M. *Proc. Natl Acad. Sci. USA* **85**, 8311–8315 (1988).
9. Overholt, E. M., Rubel, E. W. & Hyson, R. L. *J. Neurosci.* **12**, 1698–1708 (1992).
10. Reyes, A. D., Rubel, E. W. & Spain, W. J. *J. Neurosci.* **16**, 993–1007 (1996).

Obituary

Gordon Randolph Willey (1913–2002)

“Misled” and “dealt a marginal hand by my colleagues” were just two of the thoughts that ran through Gordon Willey’s mind in the spring of 1946, as he walked over the rough terrain of the Virú Valley in northern Peru. As a relative newcomer to the field of archaeology, just four years on from finishing his PhD at Columbia University, Willey fretted that, while his associates in the Virú Valley programme were getting to do the important work such as excavating stratigraphic pits or collecting ancient potsherds, he had been assigned the less significant chore of studying the distribution of archaeological sites and trying to understand their prehistoric spatial relationships and functions. Time has shown that, rather than being marginalized, Willey was on the verge of sparking methodological and interpretive innovations that would come to dominate the field of archaeology in the second half of the twentieth century. Moreover, in a few years, Willey would become the leading American archaeologist of his generation and one of the most widely respected archaeologists in the world.

Willey’s Virú Valley research soon led to the rise of a ‘settlement-pattern’ approach in archaeology. Although Willey was not the originator of settlement-pattern studies, his Virú work, and the 1953 publication that followed, convinced colleagues around the world that this particular methodology could provide new data that would in turn lead to clearer insight into the functioning of ancient cultures.

Willey’s now classic definition of settlement patterns appeared on the first page of his Virú monograph, *Prehistoric Settlement Patterns in the Virú Valley, Peru*. It is, he wrote, the study of “the way in which man disposed himself over the landscape on which he lived. It refers to dwellings, to their arrangement, and to the nature and disposition of other buildings pertaining to community life. These settlements reflect the natural environment, the level of technology on which the builders operated, and the various institutions of social interaction and control which the culture maintained. Because settlement patterns are, to a large extent, directly shaped by widely held cultural needs, they offer a strategic starting point for the functional interpretation of archaeological cultures.”

This approach became a rallying point for archaeologists who wanted to



Archaeologist of the Americas

move beyond the classificatory and typological emphases of the time, in order to understand not just the chronology of potsherds but the nature of whole cultural systems, and how and why they changed through time.

The full potential of the approach became apparent in Willey’s subsequent research in the Maya Lowlands of Belize. In 1950, Willey was appointed the first Bowditch Professor of Mexican and Central American Archaeology and Ethnology at Harvard University. Taking up this chair required him to move his research to the geographical areas stipulated in his new title. So in 1954, after brief fieldwork in Panama, Willey initiated a settlement-pattern project in the Belize River Valley, supported by the first grant for archaeological research from the newly established National Science Foundation.

Willey’s and colleagues’ settlement mapping and excavations showed how the traditional emphasis in Mayan studies — on the temples, palaces and other remains of the ruling class — had ignored the vestiges of the perishable houses and activities of Mayan peasants and others who were not members of the society’s élite. By systematically identifying these remains, Willey was able to provide a much fuller picture of ancient Mayan society than had been possible before, as well as revealing some of the pressures that this society had faced, including that of a growing population. Willey’s research rapidly stimulated other settlement-pattern work that showed, for example,

that the ancient Maya had substantial cities and complex agricultural practices to support the extensive urban zones.

As important as the settlement-pattern approach has been for archaeological research, it was not Willey’s only contribution to the field. He made significant empirical contributions to our knowledge of the prehistory of Florida and the southeastern United States in general, to our archaeological understanding of coastal Peru, and to Mayan studies with key field projects at Barton Ramie (Belize), Altar de Sacrificios and Seibal (Guatemala), and Copan (Honduras). He was also justly renowned for his broad syntheses, especially his highly influential, two-volume *An Introduction to American Archaeology* and the widely cited and discussed *Method and Theory in American Archaeology*, written with Philip Phillips. Additionally, he had a great interest in the intellectual history of archaeology, as evinced by the publication that I co-authored, *A History of American Archaeology*.

Willey’s other great scholarly talent was his ability to see the big picture. He was particularly interested in the rise and development of pre-industrial civilizations and was a strong admirer of the American anthropologist Alfred L. Kroeber. He wrote a number of notable articles on the nature of early civilizations, with special emphases on the social and ideological aspects of these complex societies.

Willey was an amiable person, widely liked and respected by his peers. He won virtually all the accolades and honours in his profession, and he trained or mentored two generations of leading Maya scholars during his teaching years at Harvard. He lectured around the world but was especially fond of the United Kingdom, spending two separate sabbatical years at the University of Cambridge and becoming a Corresponding Fellow of the British Academy.

Without doubt, Willey’s name has become the most widely known in American archaeology and his prolific publications have influenced all aspects of the discipline. Whether on a typewriter or a computer keyboard, Willey hammered out these publications with one of the fastest two-finger typing styles in the world. Countless students, colleagues and friends all over the globe will miss his illuminating ideas, breadth of knowledge and warm good humour. Jeremy A. Sabloff
Jeremy A. Sabloff is at the University of Pennsylvania Museum of Archaeology and Anthropology, 33rd and Spruce Streets, Philadelphia, Pennsylvania 19104-6324, USA.
e-mail: jsabloff@ccat.sas.upenn.edu

Parasitoid secretions provoke ant warfare

Subterfuge used by a rare wasp may be the key to an alternative type of pest control.

Insect social parasites are extreme specialists that typically use mimicry or stealth to enter ant colonies to exploit the rich, but fiercely protected, resources within their nests^{1–3}. Here we show how a parasitic wasp (parasitoid)⁴ contrives to reach its host, itself an endangered species of social parasite that lives inside the brood chambers of ant nests, by releasing semiochemicals to induce in-fighting between worker ants, locking the colony in combat and leaving it underprotected. Four of these chemicals are new to biology and have the potential to control pest species by inducing different agonistic behaviours in ants.

Ichneumon eumerus is an endangered hymenopteran parasitoid that is found in less than 20% of populations of its host, the butterfly *Maculinea rebeli*⁵, which is listed by the World Conservation Union. *I. eumerus* females deposit their eggs in *M. rebeli* larvae, and their progeny emerge 11 months later as adults from their host's pupae⁵. Both stages of *M. rebeli* inhabit colonies of the ant *Myrmica schencki*, where the larvae mimic *M. schencki* larvae in their behaviour and surface chemistry⁶. *I. eumerus* seeks its host by first detecting *M. schencki* colonies that contain *M. rebeli*, and then inducing the fighting that enables it to penetrate nests⁵.

We confirmed that *I. eumerus* uses agonistic chemicals to provoke *M. schencki* (Fig. 1). Empty *M. rebeli* pupal cases from which parasitoids had already emerged induced fierce fighting in laboratory ant colonies, as did solvent extracts from these cases painted onto dummies. In contrast, pupal cases produced no reaction after being washed in solvent (Fig. 1a, and see supplementary information).

Analysis by gas chromatography with mass spectrometry revealed that these extracts contained three alcohols and three aldehydes: 0.3 $\mu\text{g } \mu\text{l}^{-1}$ Z-9-icosen-1-ol (Z-9-C₂₀-ol); 5 $\mu\text{g } \mu\text{l}^{-1}$ Z-9-docosen-1-ol (Z-9-C₂₂-ol); 14 $\mu\text{g } \mu\text{l}^{-1}$ Z-9-tetracosen-1-ol (Z-9-C₂₄-ol); 0.3 $\mu\text{g } \mu\text{l}^{-1}$ icosanal (Z-9-C₂₀-al); 7 $\mu\text{g } \mu\text{l}^{-1}$ docosanal (Z-9-C₂₂-al); and 20 $\mu\text{g } \mu\text{l}^{-1}$ tetracosanal (Z-9-C₂₄-al). Four of these compounds are new to biology; the functions of the others (Z-9-C₂₂-al and Z-9-C₂₄-al) were previously unknown⁷.

We synthesized each chemical in the laboratory and presented it to *M. schencki* colonies on dummies, either separately or mixed in the relative concentrations found in the extract. Four chemicals provoked at least one significant response (Fig. 1b, and see supplementary information). Z-9-C₂₀-ol attracted workers to the dummy and

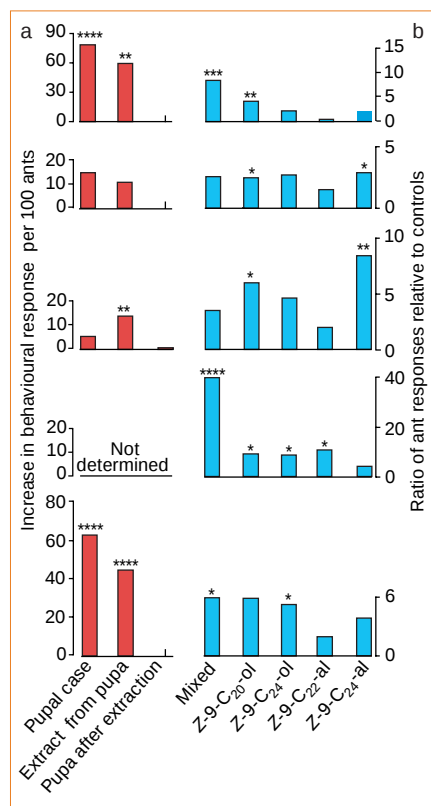


Figure 1 Responses of *Myrmica schencki* ants to chemicals produced by the parasitoid wasp *Ichneumon eumerus* and to artificially synthesized chemicals. The behaviours shown are (from top): ants run towards the source; they investigate the source; they attack it; they are repelled; they attack kin ants. **a**, Effects of biological samples produced by wasps (red), showing ants' increased behavioural responses when presented with: empty *Maculinea rebeli* butterfly pupal cases that had produced wasps, compared with cases that produced *M. rebeli*; teflon dummies dosed with hexane, ethyl acetate and methanol extracts from wasp pupae, compared with dummies dosed with solvents alone; or pupal cases after extraction. **b**, Effects of individual and mixed synthetic chemicals (blue; see text for abbreviation definitions) on ant responses relative to controls. Dummies were dosed with >99.5% pure chemicals and presented to ant colonies. The significance of results relative to controls was assessed using STATXACT (**a**) and Mann–Witney (**b**) analyses; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. For further details, see supplementary information.

encouraged them to investigate it, but, having touched it, the ants became aggressive and ran away, so that they attacked sister ants instead. Z-9-C₂₄-al promoted the initial investigation and, with Z-9-C₂₄-ol, amplified ant aggression; Z-9-C₂₂-al and Z-9-C₂₄-ol strongly repelled ants. Mixed together, these chemicals drew ants to the parasitoid, where, having become aroused to a state of high aggression, they were quickly repelled. This resulted in three to eight times more attacks ($P = 0.001$, $n = 9$)

being made on kin ants than on *I. eumerus*, the stimulator of the aggression.

The controlled propagation of alarm plays a key part in ant societies, casting a social 'immune barrier' around them and affording workers instant recognition of nestmates and discrimination against foreigners². Most of the pheromones involved in this process are volatile, short-chain hydrocarbons, which allow the rapid expansion and dissipation of alarm^{2,8,9}. Only the slave-making ant *Formica subintegra*, whose C₁₂–C₁₆ aldehydes are the most persistent and longest-chain semiochemicals known to provoke ants, has been found to use alarm in the same way as *I. eumerus* to raid other ants' nests^{2,9}. The allomones used by *I. eumerus* are more persistent still, provoking aggression for 50 days after secretion⁵.

Our bioassays induced only 56–79% of the ferocity unleashed by a living parasitoid⁵. This may be due to the small size of our test colonies — larger ones reacted more strongly ($r^2 = 92\%$, $P = 0.04$, $n = 5$), which supports our observation that the ants themselves amplify and propagate a sense of panic after contact with the wasp, creating a chain reaction of in-fighting across their society.

No other known secretion enables an individual insect to immobilize up to 80% of an ant colony. The strength and persistence of this reaction suggests that similar cocktails of long-chain hydrocarbons might provide an alternative to the use of poisons and repellents to control pest ants¹⁰. If so, this would be an unexpected benefit of studying obscure taxa¹¹ — sadly, just four unprotected European meadows now support all of the *I. eumerus* individuals found in recent decades.

J. A. Thomas*, J. J. Knapp†, T. Akino‡, S. Gerty†, S. Wakamura‡, D. J. Simcox*, J. C. Wardlaw*, G. W. Elmes*

*Centre of Ecology and Hydrology (NERC), Winfrith Technology Centre, Dorchester, Dorset DT2 8ZD, UK
e-mail: jat@ceh.ac.uk

†School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, UK

‡National Institute of Agrobiological Sciences, Ohwashi 1-2, Tsukuba 305-8634, Japan

- Elmes, G. W., Barr, B., Thomas, J. A. & Clarke, R. T. *Proc. R. Soc. Lond. B* **266**, 447–453 (1999).
- Hölldobler, B. & Wilson, E. O. *The Ants* (Springer, Berlin, 1990).
- Dettner, K. & Liepert, C. *Annu. Rev. Entomol.* **39**, 129–154 (1994).
- Godfray, H. C. J. *Parasitoids* (Princeton Univ. Press, 1994).
- Thomas, J. A. & Elmes, G. W. *Anim. Behav.* **45**, 593–602 (1993).
- Akino, T., Knapp, J. J., Thomas, J. A. & Elmes, G. W. *Proc. R. Soc. Lond. B* **266**, 1419–1426 (1999).

7. Sledge, M. F. et al. *J. Insect Physiol.* **46**, 753–761 (2000).
8. Wilson, E. O. & Bosert, W. H. *Recent Progr. Hormone Res.* **19**, 673–716 (1963).
9. Billen, J. & Morgan, E. D. in *Pheromone Communication in Social Insects* (eds Vander Meer, R. K., Breed, M. D., Winston, M. L. & Espelie, K. E.) 3–33 (Westview, Colorado, 1998).

10. Klotz, J. H., Greenberg, L., Shorey, H. H. & Williams, D. F. *J. Agri. Entomol.* **14**, 249–257 (1997).
 11. Lawton, J. H. *Ecol. Entomol.* **26**, 225–226 (2001).
- Supplementary information** accompanies this communication on Nature's website.
- Competing financial interests:** declared none.

Sports statistics

Trends and random fluctuations in athletics

Improvements in the results of athletic competitions are often considered to stem from better training and equipment, but elements of chance are always present in athletics and these also contribute. Here we distinguish between these two effects by estimating the range into which athletic records would have fallen in the absence of systematic progress and then comparing this with actual performance results. We find that only 4 out of 22 disciplines have shown a systematic improvement, and that annual best results worldwide¹ show saturation in some disciplines.

We investigated the development of sporting records by means of order statistics² that provide estimates of the maxima and minima of stochastic time series. We used the results to predict (retrospectively) how athletes' records would have evolved during the German championships^{3,4} in the absence of systematic improvements. (These can arise from an increase in the number of competition participants, a major effect during the past century.)

We assumed that the stochastic variations are the result of a strongly stationary process in which two consecutive values are independent and all values obey the same statistical distribution, $\rho(x)$. The records considered here may be either maxima (jumping and throwing distances) or minima (running times); we rendered these two types directly comparable by converting running times into average speeds, allowing us to discuss only the case of maxima. We used the values of one time interval (here, annual best results) to determine the most probable record, $x_{\max,th}$ (expectation value), and its standard deviation, σ_E , for the succeeding interval.

Assuming that $\rho(x)$ is gaussian (with mean μ and standard deviation σ), the expectation value for the record during the subsequent period of N years can be estimated as

$$x_{\max,th} = \mu + \sigma \frac{a_0 + a_1 \ln \ln(N) + a_2 (\ln \ln(N))^2}{a_2} \quad (1)$$

where the optimal coefficients for $2 \leq N \leq 100$ are (the error of the approximation of $x_{\max,th}$ for $\mu=0$ and $\sigma=1$ being $<0.06\%$ over this interval): $a_0=0.818$, $a_1=0.574$ and $a_2=0.349$ (D.G., J.G.T. and

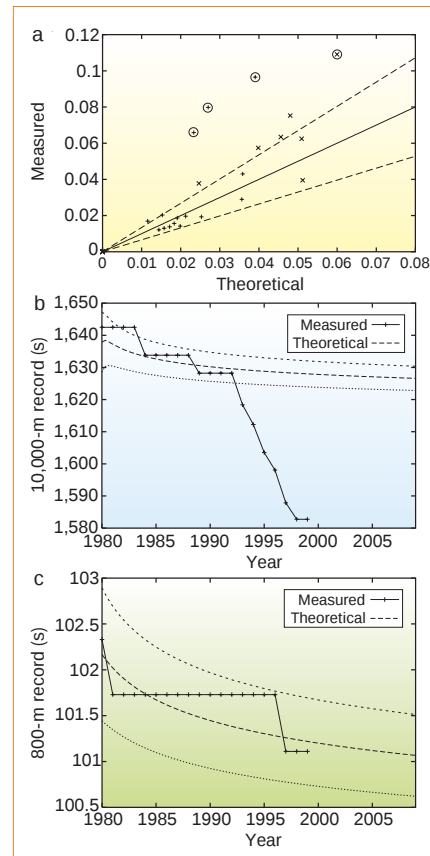


Figure 1 Trends versus random fluctuations in athletics. **a**, Comparison of actual and theoretical results for the relative differences between the best performances of the second 12-yr interval and the mean annual best result of the first 12-yr interval for the German championships. 'Theoretical' (x -axis): $(x_{\max,th}/\mu) - 1$; 'measured' (y -axis): $(x_{\max,real}/\mu) - 1$; +, running disciplines; x, jumping and throwing disciplines. The solid line $y=x$ represents perfect theoretical predictions, dashed lines (not symmetrical to the $y=x$ line) indicate $\pm$ s.d. intervals (σ_E). The greatest deviations correspond to the 110-m hurdles, 50-km and 20-km walking races, and pole vault. **b, c**, Representative forecasts of world records obtained using equation (1) and the annual best results from the period 1980–89. The time course in **b** (for the 10,000 m) shows a strong trend beginning in 1992; that in **c** for the 800 m seems to be purely stochastic.

D.S., manuscript in preparation). Analogously, we find that

$$\sigma_E = \sigma(b_0 + b_1 \ln \ln(N) + b_2 (\ln \ln(N))^2) \quad (2)$$

where $b_0=0.8023$, $b_1=-0.2751$ and $b_2=0.0020$ (approximation error $<0.15\%$). We derived these equations from formulae for maxima estimation that hold in the limit of large N (ref. 5).

We applied this method to the top performances of men in 22 disciplines during

the German championships of 1973–96. We divided the data into two groups: those for the years 1973–84 served as a reference period from which we predicted the best result for the period 1985–96. A Kolmogorov–Smirnov test showed that a gaussian function is a suitable approximation of the probability distributions of our data. The distribution should have cut-offs at zero and at some physiological limit⁶, but we found that these hypothetical cut-offs could be safely neglected.

Figure 1a compares the stochastic predictions with the actual data for all 22 disciplines. Most of the experimentally determined points are well described by the statistical analysis. Only four disciplines show a significant deviation — for these, the results achieved are better than the predictions, indicating systematic improvement. A previous analysis of world records⁷ ignored trends and considered only record-breaking events.

We also analysed a similar data set for worldwide annual best results (for the outcome of the statistical test, see supplementary information), and found that 7 out of 19 estimates do not fall into a $3\sigma_E$ interval around $x_{\max,th}$ for the time-interval pairs 1985–97 and 1990–99. Compared with the German results, stronger non-stationarities occur, indicating more systematic progress, but tend to decrease over time (D.G., J.G.T. and D.S., manuscript in preparation).

We also used the available data for 1980–89 to predict the development of world records (N here being the number of years since the pre-1980 onset of the stationary process; Fig. 1).

The method we describe here should be useful for assessing the relevance of any small signal in a noisy environment^{8,9}, and in obtaining estimations for the quality of the best (unknown) solutions of computationally complex optimization problems.

Daniel Gembris*, John G. Taylor†, Dieter Suter‡

**Institute of Medicine, Research Center Jülich GmbH, D-52425 Jülich, Germany
e-mail: d.gembris@fz-juelich.de*

†Department of Mathematics, King's College London, London WC2R 2LS, UK

‡Department of Physics, University of Dortmund, D-44221 Dortmund, Germany

1. Koponen, M. *The Track and Field Athlete of the Century According to the Statistics* (Atk-Pike, Asikkala, Finland, 2000).
2. David, H. A. *Order Statistics* (Wiley, New York, 1970).
3. Scherer, K. A. *Chronik Handbuch, Sportgeschichte in Zahlen* (Chronik, Munich, 1997).
4. Steinmetz, F. *Deutsche Meisterschaften, Band 1–4* (Hornberger, Waldschbach, Germany, 2000).
5. Galambos, J. *The Asymptotic Theory of Extreme Order Statistics* (Wiley, New York, 1978).
6. Hammond, K. A. & Diamond, J. *Nature* **386**, 457–462 (1997).
7. Tryfos, P. & Blackmore, R. J. *Am. Stat. Assoc.* **80**, 46–50 (1985).
8. Gammaitoni, L., Hänggi, P., Jung, P. & Marchesoni, F. *Rev. Mod. Phys.* **70**, 223–287 (1998).
9. Gembris, D. et al. *Magn. Reson. Med.* **43**, 259–268 (2000).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Understanding and exploiting C–H bond activation

Jay A. Labinger & John E. Bercaw

Arnold and Mabel Beckman Laboratories of Chemical Synthesis, California Institute of Technology, Pasadena, California 91125, USA

The selective transformation of ubiquitous but inert C–H bonds to other functional groups has far-reaching practical implications, ranging from more efficient strategies for fine chemical synthesis to the replacement of current petrochemical feedstocks by less expensive and more readily available alkanes. The past twenty years have seen many examples of C–H bond activation at transition-metal centres, often under remarkably mild conditions and with high selectivity. Although profitable practical applications have not yet been developed, our understanding of how these organometallic reactions occur, and what their inherent advantages and limitations for practical alkane conversion are, has progressed considerably. In fact, the recent development of promising catalytic systems highlights the potential of organometallic chemistry for useful C–H bond activation strategies that will ultimately allow us to exploit Earth's alkane resources more efficiently and cleanly.

Alkanes, or saturated hydrocarbons, are major constituents of natural gas and petroleum, but there are very few practical processes for converting them directly to more valuable products. The reason for this difficulty is alluded to by their other name, 'paraffin' (meaning 'not enough affinity'): alkanes are relatively inert. This chemical inertness arises from the constituent atoms of alkanes all being held together by strong and localized C–C and C–H bonds, so that the molecules have no empty orbitals of low energy or filled orbitals of high energy that could readily participate in a chemical reaction, as is the case with unsaturated hydrocarbons such as olefins and alkynes.

Alkanes do react at high temperatures, as encountered in combustion, but such reactions are not readily controllable and usually proceed to the thermodynamically stable and economically unattractive products, carbon dioxide and water. The currently prevalent use of alkanes in combustion applications exploits their energy content, but not their considerable potential as valuable precursors for more important and expensive chemicals. Although cracking and thermal dehydrogenation convert alkanes to valuable olefins, these processes require high temperatures and are energy intensive. Similarly, although alkanes can be induced to react by exposure to highly reactive species such as superacids or free radicals, these reactive species are usually demanding and expensive to make, and offer little control over product selectivity.

A variety of enzymes efficiently and selectively catalyse alkane oxidation at physiological temperatures and pressures, and the direct use of biological organisms for industrial alkane conversion under such benign conditions is possible in principle. But in practice, these approaches may be primarily applicable to the small-scale production of specialized chemicals, given that large-scale bioprocesses for converting alkane resources seem to be problematic¹. However, despite their practical limitations, enzymatic alkane transformations are much studied, to understand the underlying reaction mechanisms and guide the design of synthetic catalysts mimicking the function, efficiency and selectivity of their biological counterparts. Enzymatic and enzyme-mimetic alkane transformations are beyond the scope of the present review, but further information and bibliographies can be found in recent reviews^{2,3}.

Because it is difficult to achieve selective transformations under

forcing conditions, most important petrochemicals, especially oxygenates such as alcohols, aldehydes and carboxylic acids, are produced from unsaturated hydrocarbons. The primary unsaturated starting materials are olefins, in turn obtained from alkanes by fairly inefficient processes. Milder and better-controlled direct conversions of alkanes into olefins may thus offer large economic benefits. Alkane chemistry could also improve the utilization of methane, the principal component of natural gas. Many of the world's established natural gas resource locations are remote, in sites where there is little or no local demand, such as the north slope of Alaska. Exploitation of such resources is impeded by the high cost of both gas transportation and current methods for converting hydrocarbon gas into more readily transportable liquid⁴. The available conversion methods are indirect, involving production of synthesis gas (carbon monoxide and hydrogen), followed by conversion to the desired product. Developing efficient strategies for the direct conversion of methane to methanol or other liquid fuels or chemicals could thus significantly improve methane utilization.

The controlled activation of small, relatively inert molecules has been a ubiquitous theme in transition metal chemistry since its renaissance began in the late 1950s. Practitioners of this art are thus well placed to address the alkane activation problem. To date, binding to metal centres has resulted in altered and/or enhanced reactivity in a variety of molecules, through associated changes in the relative energies of their orbitals or their polarity. Olefins and carbon monoxide, for example, are rendered more susceptible to nucleophilic attack upon coordination to a metal centre⁵, whereas coordinated dioxygen frequently reacts with electrophiles⁶. Even dinitrogen, the classic inert small molecule, can be thus activated and induced to partake in chemical transformations⁷. In principle, it should be possible to similarly activate the inert C–H bonds of alkanes, but the examples just mentioned all involve small molecules with lone electron pairs and/or π orbitals that can interact with empty orbitals of the metal centres; alkanes possess neither of these.

Indeed, in most pre-1980s examples of C–H bonds reacting at transition-metal centres, reactions proceed successfully only if they are assisted through participation of the π orbitals of aromatic C–H bonds⁸ (equation (1)) or involve intramolecular reactions, where the group containing the C–H bond to be activated is already attached to the metal⁹ (equation (2)). Often both effects contribute¹⁰ (equation (3)).

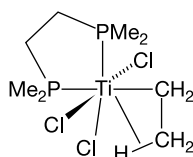
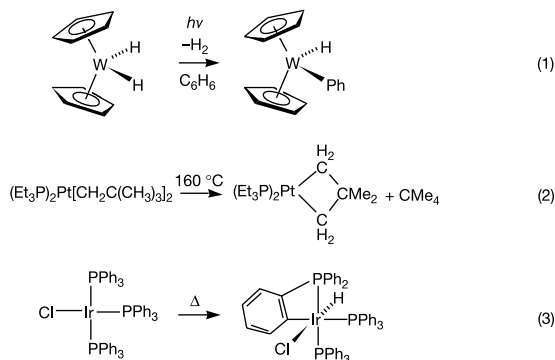


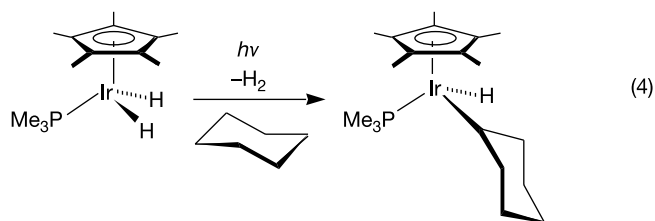
Figure 1 An example of agostic bonding. The electron pair of a β -C-H bond from a coordinated ethyl group interacts with a vacant orbital on Ti (ref. 13).



where $h\nu$ implies light energy and Δ implies heat energy. Functional groups are abbreviated as follows: Ph, phenyl; Et, ethyl; Me, methyl.

If such assistance were in fact required, activation of aliphatic C-H bonds at transition-metal centres would not be a good strategy for catalytic alkane conversion. However, during the 1970s, examples of significant interaction between a metal centre and an alkane C-H bond suggested that this difficulty might have been overestimated. For example, some metal complexes were known to catalyse alkane reactions such as hydrogen-deuterium (H-D) exchange or oxidation under relatively mild conditions¹¹. Moreover, several cases of so-called agostic bonding were discovered¹², wherein a C-H bond of an already coordinated ligand forms an additional bond to the metal (see Fig. 1 for a typical example¹³).

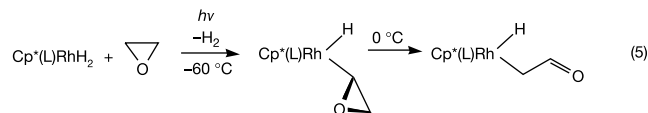
These observations suggested that alkane activation should be possible. In fact, chemists realized that previous failures to achieve it might not have been due to unreactivity at all: C-H activation might take place, but yield an organometallic product that is thermodynamically unstable and thus undetectable¹⁴. This interpretation proved to be correct. Using systems designed to produce stable products, two groups demonstrated intermolecular alkane activation in 1982 (refs 15, 16) (one reaction is shown in equation (4)), and a large number of other examples quickly followed. (Ref. 17 provides a thorough survey of these developments and further references.)



Challenges for practical alkane conversion

Although the accomplishments of C-H activation mentioned so far afford new organometallic derivatives, the development of practical alkane conversion strategies poses further, and perhaps more

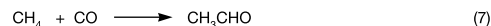
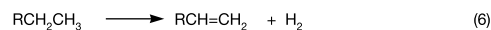
difficult, challenges. One is activity, as only reactions proceeding at an adequate rate will be useful. Another is selectivity, which is challenging for two reasons. First, the desired product must not have any functional group that reacts more readily with the metal centre than the alkane starting material. In this regard, organometallic activation appears promising: as strikingly illustrated by the example in equation (5) an 'inert' C-H bond is activated by a metal-centred reagent, even though the nearby epoxide linkage is far more reactive under most conditions¹⁸.



where Cp^* is pentamethylcyclopentadienyl.

A second and more general problem with achieving selectivity arises from the relative reactivity of different C-H bonds. Most existing transformations of alkane C-H bonds involve oxidation, whether in the gas phase, over metal-oxide catalysts at high temperature, or in solution. These reactions are dominated by radical pathways involving hydrogen-atom removal; the reactivity of such processes tends to correlate with C-H bond strength¹⁹. The terminal positions of alkanes will therefore be the least reactive, yet terminally functionalized alkanes are often the most desirable products, as exemplified by the role of terminal alcohols in detergents. Furthermore, C-H bonds adjacent to functional groups in oxidized products, such as alcohols and aldehydes, are more reactive than the C-H bond of the original alkane; transformations of alkanes proceeding via radical pathways will thus be subject to an inherent upper limit on the product yield that can be obtained. The relative reactivity for activating different C-H bonds is hence a crucial issue.

Finally, thermodynamics presents perhaps the highest hurdle to a net overall alkane conversion. Transformations such as dehydrogenation (equation (6)) and carbonylation (equation (7)) should readily be accomplished at centres that activate C-H bonds, but both of these reactions are thermodynamically uphill anywhere near room temperature. The addition of a C-H bond across an olefin, as in (equation (8)), is favourable, but serves only to produce another alkane; such a transformation will generally be of little use. (In cases where C-H activation is used to add arenes across a C-C multiple bond²⁰⁻²², the process can potentially yield valuable substituted aromatic products.)



where R and R' indicate different alkyl groups.

The partial oxidation of alkanes to alcohols, aldehydes, acids and other oxygenates is, in contrast to the above reactions, thermodynamically favoured and also leads to highly desirable products. However, many of the known C-H activating centres are highly sensitive to O_2 and other oxidants.

Can the demands of thermodynamics and chemical compatibility, along with activity and selectivity requirements, all be satisfied simultaneously? To answer this question, we now consider the types of C-H activation reactions at metal centres and their underlying mechanisms.

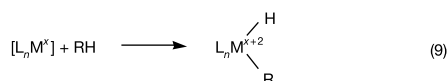
Classification of reactions

The scope of this review is limited to C-H activation at metal centres, so we restrict our discussion to reactions that have clearly been shown to involve bond formation between a metal and an alkane carbon. The reactions are conveniently classified according

to their overall stoichiometry. There are four classes involving the formation of stable organometallic species, and two of these occur quite commonly, while the other two are rare. The fifth class of activation reactions involves transient generation of organometallic species as reaction intermediates.

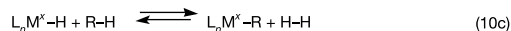
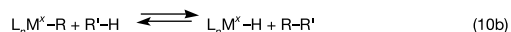
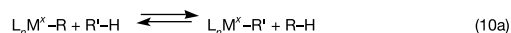
Oxidative addition

Oxidative addition reactions are typical for electron-rich, low-valent complexes of the 'late' transition metals found towards the right side of the periodic table—Re, Fe, Ru, Os, Rh, Ir, Pt. In this reaction type, illustrated in equation (9), the reactive species $[L_nM^x]$ is coordinatively unsaturated and hence almost always unstable; it is therefore generated *in situ* by thermal or photochemical decomposition of a suitable precursor. A typical precursor is $(\eta^5-C_5Me_5)-(PMe_3)Ir^{III}H_2$ (see equation (4)), which loses H_2 under photo-irradiation to give, as reactive species $[L_nM^x]$, the (unobserved) intermediate $[(\eta^5-C_5Me_5)(PMe_3)Ir^I]$ (ref. 15).



Sigma-bond metathesis

Alkyl or hydride complexes of 'early' transition metals with d^0 electronic configurations may undergo the reversible reaction of equation (10a) (refs 23, 24). These metals are most commonly from group 3 of the periodic table (Sc, lanthanides and actinides), but some examples involving metals of groups 4 and 5 are also known²⁵.

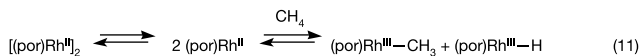


Most commonly R and R' are both alkyl groups, in which case the reaction only amounts to the interchange of alkyl fragments, rather than net alkane activation. The making and breaking of C–C single bonds through the alternative exchange (equation (10b)) is unfortunately not observed, whereas some examples of interconversion of dihydrogen and metal alkyl with alkane and metal hydride (equation (10c)) are known^{26,27}.

Metalloradical activation

Rhodium(II) porphyrin complexes exist in a monomer–dimer equilibrium and can reversibly break alkane C–H bonds, with attachment of the two fragments to two separate rhodium(II) porphyrin complexes (see equation (11)). Methane is the most reactive hydrocarbon for this class of reaction. Regarding the

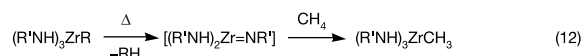
underlying mechanism, the involvement of free alkyl radicals, generated through the abstraction of a hydrogen atom from methane by the Rh centre, might at first seem attractive. But the low Rh–H bond strength (around 60 kcal mol^{−1} compared to the 105 kcal mol^{−1} C–H bond of methane) suggests that such a step would be prohibitively endothermic, in agreement with inferences from experimental observations^{28,29}.



where (por) is porphyrin.

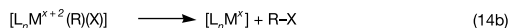
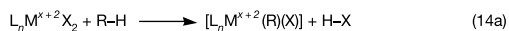
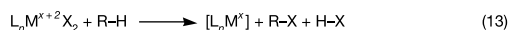
1,2-addition

1,2-addition reactions involve the addition of an alkane to a metal–nonmetal double bond. An example of methane addition across a Zr–N double bond (equation (12)) has been reported³⁰, while a related system activates benzene but not methane³¹. Although alkane additions across M = N and M = C double bonds of other early and middle transition-metal centres^{32–34} are also known, the scope of this type of reaction and its potential for alkane functionalization remain unclear.



Electrophilic activation

Certain reactions that lead directly to functionalized alkanes, rather than observable organometallic species (although their participation as intermediates is strongly indicated), have been classified as electrophilic activation reactions. This type of reaction is illustrated in equation (13) where $[M^{x+2}]$ is a late- or a post-transition metal (Pd^{2+} , Pt^{2+} and/or Pt^{4+} , Hg^{2+} , Tl^{3+}), usually in a strongly polar medium such as water or an anhydrous strong acid³⁵. The presumed organometallic intermediate $[L_nM^{x+2}(R)(X)]$ involved in the transformation would be formed as indicated in equation (14a), that is, through substitution of a metal for a proton—hence the term 'electrophilic' activation. Equation (14b) shows a possible route from the intermediate to the product.



Alkane activation mechanisms

While most of our understanding of alkane activation mechanisms has been obtained from studies of oxidative addition reactions, it

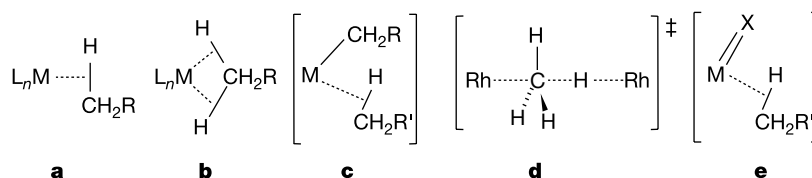


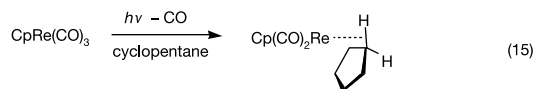
Figure 2 Possible structures of sigma complexes. **a**, The simplest structure, analogous to that observed in agostic complexes. **b**, A 'dihapto' structure with two coordinated C–H bonds, suggested by theoretical calculations. **c**, A possible intermediate in a sigma-bond metathesis reaction. **d**, The proposed transition state (‡) for metalloradical C–H activation. **e**, A possible intermediate in a 1,2-addition reaction.

seems likely that all C–H activations begin with the formation of an intermediate alkane complex; the subsequent C–H bond cleavage step will then differ from one class to another.

The initial interaction

The existence of stable agostic complexes with C–H bonds coordinated to a metal centre (see above) suggested that C–H bond activation might involve the initial formation of a so-called ‘sigma complex’, in which a metal centre interacts with the electron pair forming the C–H σ -bond³⁶. Direct analogy with agostic species suggests that of the possible sigma-complex structures shown in Fig. 2, structure **a** is the most probable one. However, theoretical calculations and mechanistic considerations suggest a range of other possible structures as well³⁶. No stable sigma complexes have yet been isolated, but a substantial body of evidence suggests that they do exist.

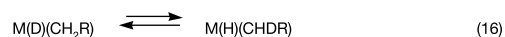
For example, vibrational spectroscopic signatures detected both in frozen matrices and in solution have been assigned to sigma complexes generated by flash photolysis of a metal carbonyl in the presence of an alkane. In these experiments, typically only the C–O stretching bands are detected and the assignment is thus not conclusive. However, nuclear magnetic resonance (NMR) examination of the longest-lived of these suspected sigma complexes, $\text{CpRe}(\text{CO})_2(\text{RH})$ (see also equation (15)), revealed a high-field signal characteristic of a C–H–M interaction³⁷.



where Cp is $(\eta^5\text{-C}_5\text{H}_5)$.

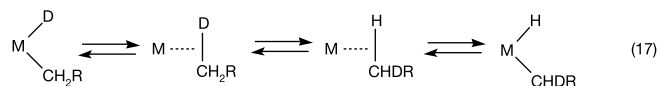
The $\text{CpRe}(\text{CO})_2(\text{RH})$ sigma complex does not proceed to form a stable C–H oxidative addition product, but vibrational spectroscopic evidence for the transient presence of sigma complexes has also been obtained in true C–H activating systems. For example, photolysis of $\text{Cp}^*\text{Rh}(\text{CO})_2$ in a mixture of alkane and liquid inert gas (xenon or krypton) yields two successive metastable intermediates. The first intermediate is an inert gas complex, $\text{Cp}^*(\text{CO})\text{Rh}(\text{Kr}$ or $\text{Xe})$, followed by formation of the alkane complex $\text{Cp}^*(\text{CO})\text{Rh}(\text{RH})$, which subsequently undergoes C–H bond cleavage to give the final product (Fig. 3)^{38,39}.

Isotope exchange has provided further evidence for the role of metastable sigma complexes. The exchange process in equation (16), for example, has been observed in over a dozen systems, and in one case is fast even on the NMR timescale⁴⁰.



Because it has been shown that complete loss of alkane from the

metal centre and subsequent re-addition does not take place³⁶, the exchange must proceed through an intermediate alkane complex, as suggested in equation (17).

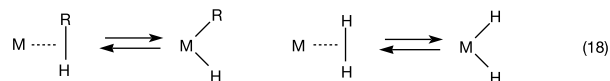


The reductive elimination of alkanes often also exhibits inverse kinetic isotope effects; that is, the reaction is faster for $\text{M}(\text{D})(\text{R})$ than for $\text{M}(\text{H})(\text{R})$. Such observations suggest that the reductive elimination processes proceed through a late transition state that involves dissociation of the alkane sigma complex after the C–H bond (stronger than the M–H bond) has already formed.

Overall, even though alkane sigma complexes may remain elusive, compelling evidence for their existence has been obtained. In fact, the C–H–M interaction in these complexes is fairly strong (of the order of about 5–10 kcal mol^{−1})³⁶, and sigma complexes seem to be important intermediates (not merely transition states) in C–H activation reactions.

Cleaving the C–H bond

Once an alkane complex has formed, the coordinated C–H bond is cleaved to yield the product. In the case of oxidative addition reactions (class 1), this step seems formally analogous to the interconversion between dihydrides and dihydrogen complexes, which is often extremely facile⁴¹. (This analogy is illustrated in equation (18)). In terms of electron counting—the organometallic chemist’s favourite predictive tool—there is no obvious barrier to interconverting $\text{M}(\text{RH})$ with $\text{M}(\text{R})(\text{H})$. However, as illustrated in Fig. 4, the bond cleavage step may not be straightforward because ligand dissociation often precedes ligand interconversion, even when not required by electron count considerations⁴².



For the other classes of C–H activation reactions, direct information on the C–H bond cleavage mechanism is relatively sparse. But theoretical and experimental⁴³ findings suggest that sigma bond metathesis reactions (class 2) and 1,2 addition reactions (class 4) proceed through an intermediate sigma complex (possible structures are shown in Fig. 2 as **c** and **e** respectively), which facilitate C–H cleavage by a 4-centre transition state. In the case of metalloradical activation of methane by $\text{Rh}(\text{porphyrin})$ complexes (class 3), the activation of a C–H bond seems to involve coordination to two metals, one to the C atom, the other to the H atom²⁹

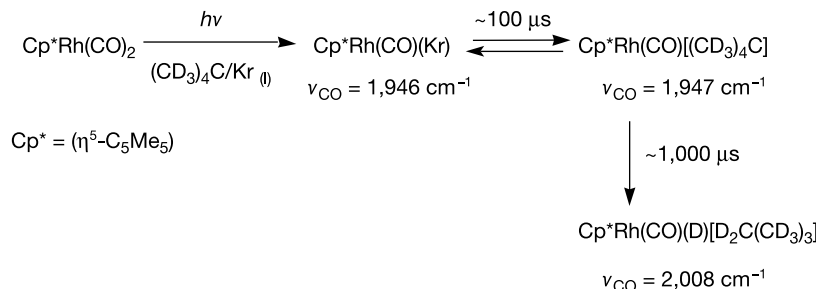


Figure 3 Pathway for oxidative addition of an alkane. The pathway has been inferred from time-resolved infrared spectroscopy observations^{38,39}. After photolytic expulsion of a carbonyl ligand from the organometallic complex, equilibrium between an inert gas complex and an alkane complex is rapidly established; the latter undergoes a slower C–H bond cleavage to generate the final product. ν_{CO} , wavenumber.

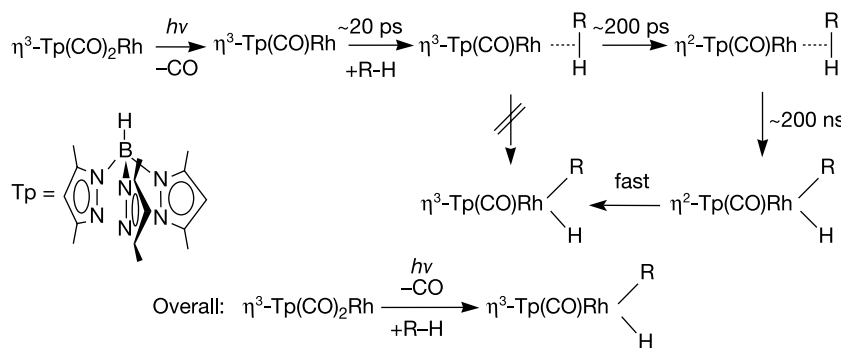
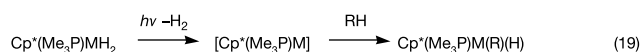


Figure 4 Detailed mechanism of oxidative addition. The alkane complex formed rapidly after CO elimination does not lead directly to C–H bond cleavage. Instead one of the three coordinated nitrogen centres of the Tp ligand dissociates on the picosecond timescale, generating a new intermediate. The subsequent C–H bond-breaking step occurs on the nanosecond timescale, followed by rapid recoordination of the third nitrogen to form the final product⁴².

(species **d** in Fig. 2 shows the proposed structure of this sigma complex).

Reaction selectivity

Transition-metal centres preferentially activate terminal C–H bonds, in contrast to the selectivity seen in radical C–H cleavages. For example, the reaction shown in equation (19) gives 100% primary alkyl for $M = \text{Rh}$ (ref. 44), and 70–80% (depending on alkyl chain length) for $M = \text{Ir}$ (ref. 45).



The different reactivities of different alkane species undergoing oxidative addition to a photolytically activated rhodium carbonyl complex (the reaction shown in Fig. 3) suggest that less substituted positions will be more reactive in the C–H bond activation step. But the opposite trend seems to hold, at least in some cases, for the sigma-complex formation step⁴⁶. Thus the overall selectivity of alkane activation reactions may well exhibit a complex dependence

on competing factors, with the preferences, relative rates and reversibilities of individual steps all having the potential to exert an influence.

Towards practical applications

Mechanistic insight into the fundamental processes and interactions controlling alkane activation and conversion will guide the development of catalysts that are compatible with thermodynamic and economic constraints, yet exhibit the activity and selectivity necessary for practical applications. Although systems suitable for practical use have not yet been developed, several promising examples point the way towards approaches that may well prove successful.

Oxidation via electrophilic activation

Net alkane oxidation is often achieved through electrophilic activation (reaction class 5 above) by oxidation-tolerant complexes. The earliest example is the platinum chemistry discovered by Shilov and co-workers⁴⁷ (equation (20)).

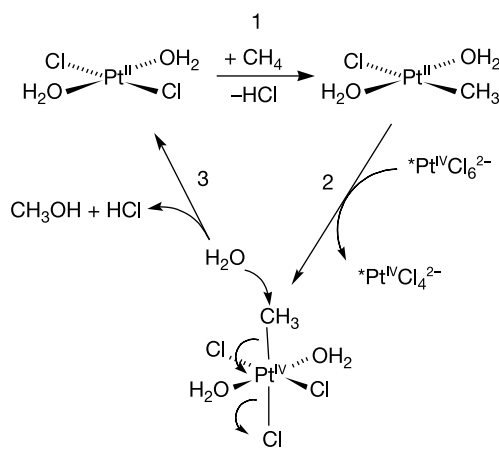
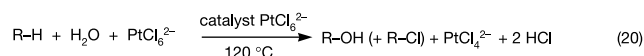


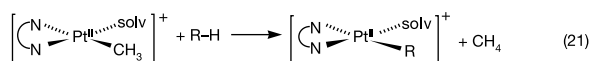
Figure 5 Mechanism of methane oxidation in the Shilov system⁴⁸. In step 1, Pt(II) effectively replaces a proton to yield a methyl–Pt(II) complex. In step 2, $[\text{PtCl}_6]^{2-}$ oxidizes the complex by electron transfer to a methyl–Pt(IV) species, which undergoes nucleophilic attack at the C–Pt bond in step 3, liberating methanol and regenerating the Pt(II) catalyst.

The reaction sequence for this system (Fig. 5) shows that although the overall conversion is catalytic in Pt(II), it requires stoichiometric amounts of Pt(IV) and is thus impractical. Moreover, the catalytic species are not indefinitely stable in solution under reaction conditions and eventually precipitate as metallic platinum, owing in part to the fact that the Pt(II)–Pt(0) redox potential is very close to that of the Pt(IV)–Pt(II) couple (ref. 48 and references cited therein).

Despite its practical limitations, this system and model analogues have revealed interesting and potentially useful features, including nonradical selectivity patterns similar to those found in well-characterized C–H activation systems. However, selectivities in this system are generally not quite as high as in the model analogues. For example, pentane is converted to *n*-pentyl chloride with 56% terminal selectivity⁴⁹; as noted earlier, oxidative addition reactions typically achieve terminal selectivities of 70–100%. Ethanol oxidation by platinum complexes exhibits a particularly unusual selectivity pattern, with functionalization at the methyl group yielding significant amounts of ethylene glycol and 2-chloroethanol⁵⁰; virtually all other known oxidation methods oxidize the alcohol function only. Ethylene glycol can even be obtained from ethane by this route⁵¹.

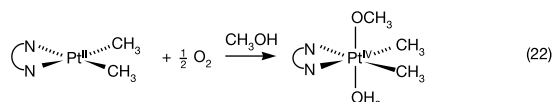
Another promising feature of the platinum chemistry is the fact that the key oxidation step (the second step in Fig. 5) involves electron transfer rather than alkyl-group transfer⁴⁸. If oxidation had required alkyl-group transfer, Pt(IV) would seem an obligatory stoichiometric oxidant for the conversion, thus rendering the overall process prohibitively expensive. Given that oxidation is achieved through electron transfer, Pt(IV) could in principle be replaced by a cheaper and more practical oxidant. Indeed, alkyl groups made soluble in water as sulphonic acids have been oxidized to alcohols using either electrochemical oxidation⁵², or O₂ in combination with heteropoly acids⁵³ or copper salts⁵⁴. In these modified processes, the involvement of platinum is catalytic, but so far only a modest number of turnovers has been achieved.

Similarly promising is successful C–H activation under very mild conditions by ligand-substituted Pt(II) complexes^{55,56}, generally of the form [(N–N)Pt(CH₃)(solv)]⁺, where N–N is a bidentate nitrogen-centred ligand and ‘solv’ is a weakly coordinating solvent (equation (21)).



where solv is NC₅F₅ or CF₃CH₂OH.

Moreover, platinum(II) complexes closely related to those formed in equation (21) have been successfully oxidized by O₂, again under mild conditions (equation (22))⁵⁷.



These findings, in combination, suggest that it might be possible to develop a Pt(II) complex that is stabilized against reduction to Pt metal by strongly bound ligands, and capable of activating alkanes and catalysing their oxidation to alcohols by O₂. Figure 6 outlines

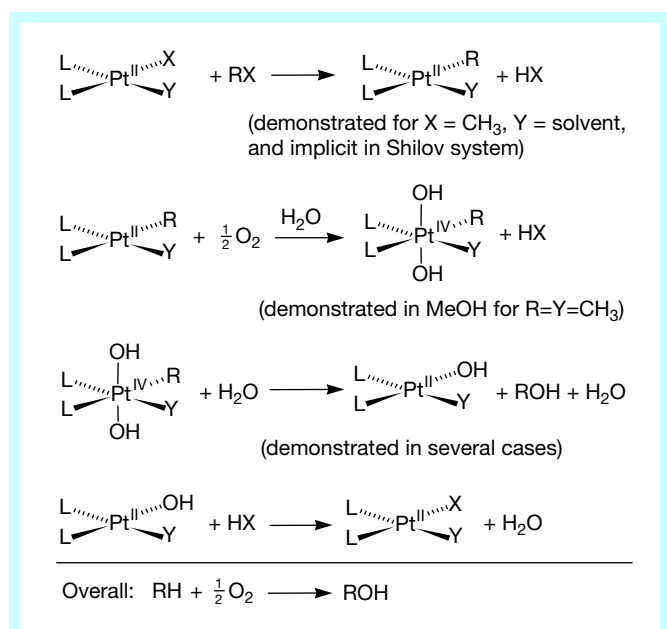


Figure 6 Hypothetical sequence for alkane oxidation to alcohol by O₂ and catalysed by a Pt(II) complex. The four individual steps—electrophilic activation of alkane at Pt(II), oxidation of RPt(II) to RPt(IV) by O₂, nucleophilic cleavage of the R–Pt bond to form ROH, and ligand exchange to regenerate the original Pt(II) species—have all been demonstrated in one or more cases, but a complex that effects all four steps has not yet been developed.

the sequence of reaction steps involved in such a catalytic alkane oxidation scheme. Although each individual step in this scheme has been realized with at least one platinum complex, a single complex accomplishing all four steps and capable of closing this catalytic cycle has not yet been found.

Oxidation of methane by sulphuric acid

From a practical perspective, perhaps the most impressive accomplishment to date is the oxidation of methane to methyl bisulphate by sulphuric acid, catalysed by a Pt(II) complex⁵⁸. The reaction appears to be an example of electrophilic functionalization, and the proposed mechanism (Fig. 7a) is closely related to the chemistry discussed above in ‘Oxidation via electrophilic activation’. In addition, the organometallic complex used in this system is remarkably stable under the reaction conditions used: the ligand is not oxidized, and there is no platinum metal formation. The absence of metal precipitation is a thermodynamic effect: the combination of the chosen ligand and hot sulphuric acid can dissolve platinum metal. A second improvement is the fact that product can be obtained with up to 90% selectivity. This feature is in part due to the protective power of the bisulphate group: to achieve the observed yield, the relative reactivity of methane versus methyl bisulphate must be of the order of 100:1.

However, the high selectivity achieved comes at a price: the ‘protected’ product is of little direct use and would need to be separately converted to a more useful compound, such as methanol, using a scheme such as that shown in Fig. 7b. At least at present, such an integrated multistep process seems not to be economically competitive with the currently used technology, the indirect conversion of methane to methanol via synthesis gas.

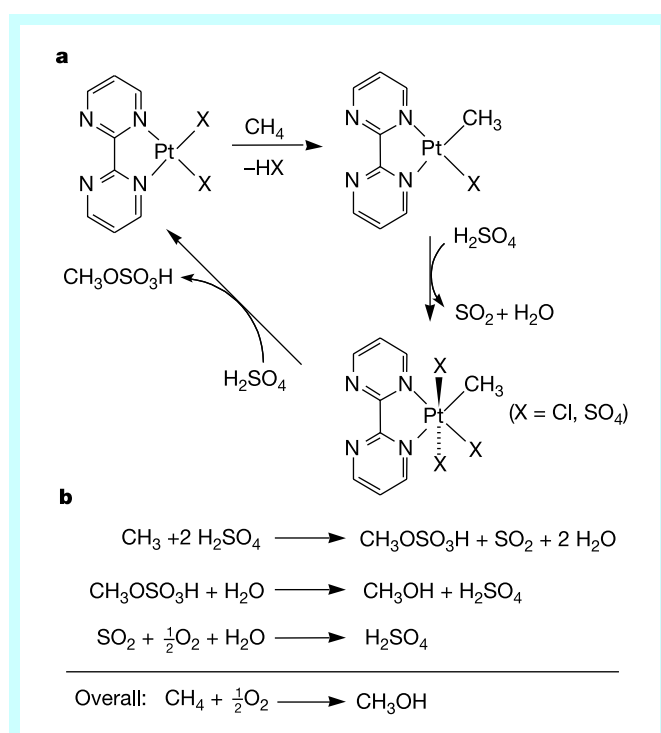
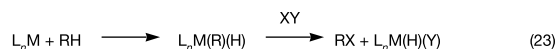


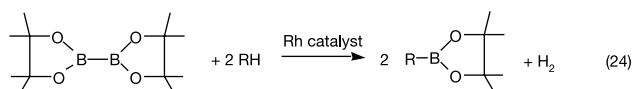
Figure 7 The Catalytica system for methane oxidation⁵⁸. **a**, Proposed mechanism for oxidation of methane by sulphuric acid, catalysed by Pt(II), by analogy to the mechanism of Fig. 5. **b**, A possible three-step sequence for methane conversion, consisting of (1) oxidation of methane to methyl bisulphate as in **a**; (2) hydrolysis of methyl bisulphate to methanol and sulphuric acid; (3) oxidation of SO₂ produced in step (1) back to sulphuric acid. The overall process accomplishes the selective oxidation of methane to methanol by O₂. Although all three steps are well established, it is not clear whether the process is economically viable.

Catalytic borylation of alkanes

If a metal-alkyl complex obtained through C–H activation could be transformed according to equation (23) into a functionalized alkane RX and a new ‘recyclable’ metal complex, in which the alkyl ligand has been replaced by a ligand Y (this might be an oxygen-centred ligand, or a halide), a useful process could be designed. But typically the M–Y bond is very stable, thus precluding closure of a catalytic cycle.



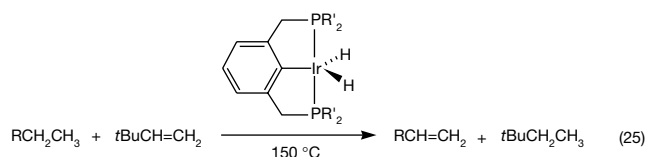
However, a first example of successful evolution of stoichiometric to catalytic functionalization using transition-metal boryl complexes has been reported. Initially, photolysis of the metal boryl complex $Cp^*W(CO)_3(Bcat')$ (where cat' is 3,5-dimethylcatecholate) in the presence of alkanes RH was shown to produce $RBcat'$ in good yield, and with complete selectivity for terminal positions⁵⁹. The same products were then generated catalytically, using a dimeric boron reagent such as $(Bpin)_2$ (where pin is pinacolate) and $Cp^*Re(CO)_3$ as catalyst⁶⁰. Photolysis is still needed in this system, but the photochemical energy input is not inherently required to overcome unfavourable thermodynamics of the overall reaction. Instead, the need simply reflects the difficulty in expelling a carbonyl ligand from the metal complex to make a vacant site available. Use of a more labile complex as catalyst should thus allow the transformation under strictly thermal activation. Indeed, Rh(I) olefin complexes such as $Cp^*Rh(C_2H_4)_2$ or $Cp^*Rh(\eta^4-C_6Me_6)$ catalyse the reaction shown in equation (24), producing $RBpin$ in yields of up to 90% (based on $(Bpin)_2$) and with the same total terminal selectivity observed in the stoichiometric versions of the reaction⁶¹. A related Ir-catalysed borylation of benzene has now also been reported⁶².



The high selectivity and the versatility of the alkylboranes precursors offer many intriguing application possibilities. However, the high cost of the borane reagents might preclude their use in large-scale processes; in contrast, smaller-scale uses certainly seem feasible.

Catalytic alkane dehydrogenation

As noted before, alkane dehydrogenation at low temperatures is thermodynamically uphill, but this problem may be overcome by providing a hydrogen acceptor to shift the equilibrium of the reaction⁶³. Examples of such catalytic ‘transfer dehydrogenation’ have been reported, but most of the catalysts used exhibit poor thermal stability at the temperatures needed to achieve reasonable reaction rates. A new class of catalyst, so-called ‘pincer’ complexes, exhibit high thermal stability, making them useful for this transformation (equation (25))⁶⁴.



The most effective ‘pincer’ catalyst, with $R' = CHMe_2$, catalyses the transfer dehydrogenation of cyclic and linear alkanes at 150 °C using hydrogen acceptors such as norbornene or *t*-butylethylene, at rates of up to 3–4 turnovers min^{-1} (ref. 65). Of particular practical

interest is the initially high selectivity in converting linear alkanes to terminal alkenes: at early stages, α -olefin selectivities of more than 90% can be obtained. As the reaction proceeds, however, competing isomerization causes a gradual shift in product selectivity to the thermodynamically favoured internal olefins.

The high thermal stability of the ‘pincer’ complexes even permits dehydrogenation without use of an acceptor: refluxing the catalyst in a high-boiling alkane such as cyclooctane (boiling point 151 °C) or cyclodecane (boiling point 201 °C) allows the liberated H_2 to escape and thus shifts the equilibrium of the reaction. This procedure gives the corresponding cycloalkenes; in the case of cyclodecane, product is formed at rates as high as nearly 8 turnovers min^{-1} (ref. 66). However, increasing concentrations of olefin product gradually inhibit the reaction. As a consequence, overall reaction rates are lower than those of the corresponding transfer dehydrogenations and isomerization becomes increasingly competitive and even dominant, thereby suppressing selectivity for terminal dehydrogenation.

The challenges ahead

The past twenty years have seen a substantial shift of focus and thinking in the field of C–H bond activation at metal centres. Initially, much effort was aimed at elucidating whether alkane activation at metal centres was at all possible, resulting in the realization that this process is, in fact, easier than had been anticipated. Attention then turned to investigating how these reactions take place. Although much remains to be learned about the processes and factors controlling the activity and selectivity of catalytic alkane activation, considerable mechanistic insight has been gained, particularly for oxidative addition reactions. The main difficulty ahead now lies with using and increasing this knowledge in order to develop practical processes that exploit alkane transformation chemistry. Although it remains challenging to reconcile the constraints inherent in alkane chemistry with thermodynamic, economic and engineering requirements when trying to develop such processes, the achievements to date promise that ongoing research in this important field will eventually allow us to harness alkanes more directly, efficiently and cleanly. □

- Duetz, W. A., van Beilen, J. B. & Witholt, B. Using proteins in their natural environment: potential and limitations of microbial whole-cell hydroxylations in applied biocatalysis. *Curr. Opin. Biotech.* **12**, 419–425 (2001).
- Ortiz de Montellano, P. R. (ed.) *Cytochrome P450: Structure, Mechanism and Biochemistry* 2nd edn (Plenum, New York, 1995).
- Brazeau, B. J. & Lipscomb, J. D. in *Enzyme-Catalyzed Electron and Radical Transfer* (ed. Holzenburg, A. Scrutton, N. S.) 233–277 (Kluwer, New York, 2000).
- Gradassi, M. J. & Green, N. W. Economics of natural gas conversion processes. *Fuel Proc. Technol.* **42**, 65–83 (1995).
- Collman, J. P., Hegedus, L. S., Norton, J. R. & Finke, R. G. *Principles and Applications of Organotransition Metal Chemistry* Ch. 7, 2nd edn (Univ. Science, Mill Valley, 1987).
- Sheldon, R. A. & Kochi, J. K. *Metal-Catalyzed Oxidations of Organic Compounds* Ch. 4 (Academic, New York, 1981).
- Fryzuk, M. D. & Johnson, S. A. The continuing story of dinitrogen activation. *Coord. Chem. Rev.* **200**, 379–409 (2000).
- Green, M. L. H. & Knowles, P. J. Formation of a tungsten phenyl hydride derivative from benzene. *J. Chem. Soc. Chem. Commun.* 1677 (1970).
- Foley, P. & Whitesides, G. M. Thermal generation of bis(triethylphosphine)-3,3-dimethylplatinacyclobutane from dincopentylbis(triethylphosphine)platinum(II). *J. Am. Chem. Soc.* **101**, 2732–2733 (1979).
- Bennett, M. A. & Milner, D. L. Chlorotris(triphenylphosphine)iridium(I): An example of hydrogen transfer to a metal from a coordinated ligand. *J. Chem. Soc. Chem. Commun.* 581–582 (1967).
- Shilov, A. E. & Shteinman, A. A. Activation of saturated hydrocarbons by metal complexes in solution. *Coord. Chem. Rev.* **24**, 97–143 (1977).
- Brookhart, M. & Green, M. L. H. Carbon-hydrogen-transition metal bonds. *J. Organometall. Chem.* **250**, 395–408 (1983).
- Dawoodi, Z., Green, M. L. H., Mtetwa, V. S. B. & Prout, K. Evidence for a direct bonding interaction between titanium and a beta-C–H moiety in a titanium-ethyl compound: X-ray crystal-structure of $[Ti(Me_2PCH_2CH_2PMe_2)EtCl_3]$. *J. Chem. Soc. Chem. Commun.* 802–803 (1982).
- Halpern, J. Activation of carbon hydrogen-bonds by metal complexes: mechanistic, kinetic and thermodynamic considerations. *Inorg. Chim. Acta* **100**, 41–48 (1985).
- Janowicz, A. H. & Bergman, R. G. C–H activation in completely saturated hydrocarbons: direct observation of $M + R-H \rightarrow M(R)(H)$. *J. Am. Chem. Soc.* **104**, 352–354 (1982).
- Hoyano, J. K. & Graham, W. A. G. Oxidative addition of the carbon hydrogen-bonds of neopentane and cyclohexane to a photochemically generated iridium(I) complex. *J. Am. Chem. Soc.* **104**, 3723–3725 (1982).

17. Shilov, A. E. & Shul'pin, G. B. *Activation and Catalytic Reactions of Saturated Hydrocarbons in the Presence of Metal Complexes* (Kluwer Academic, Dordrecht, 2000).
18. Wu, J. & Bergman, R. G. Conversion of epoxides to rhodium enolates: direct evidence for a mechanism involving initial C-H activation. *J. Am. Chem. Soc.* **111**, 7628–7630 (1989).
19. Russell, G. A. in *Free Radicals* (ed. Kochi, J. K.) 275–331 (Wiley, New York, 1973).
20. Jordan, R. F. & Taylor, D. F. Zirconium-catalyzed coupling of propene and alpha-picoline. *J. Am. Chem. Soc.* **111**, 778–779 (1989).
21. Jia, C. *et al.* Efficient activation of aromatic C–H bonds for addition to C–C multiple bonds. *Science* **287**, 1992–1995 (2000).
22. Matsumoto, T., Taube, D., Periana, R. A., Taube, H. & Yoshida, H. Anti-Markovnikov olefin arylation catalyzed by an iridium complex. *J. Am. Chem. Soc.* **122**, 7414–7415 (2000).
23. Watson, P. L. Methane exchange reactions of lanthanide and early transition metal methyl complexes. *J. Am. Chem. Soc.* **105**, 6491–6493 (1983).
24. Thompson, M. E. *et al.* "Sigma bond metathesis" for C–H bonds of hydrocarbons and Sc–R (R = H, alkyl, aryl) bonds of permethylcandocene derivatives. Evidence for the noninvolvement of the pi system in electrophilic activation of aromatic and vinylic C–H bonds. *J. Am. Chem. Soc.* **109**, 203–219 (1987).
25. Rothwell, I. P. Cyclometalation chemistry of aryl oxide ligation. *Acc. Chem. Res.* **21**, 153–159 (1988).
26. Vidal, V., Théolier, A., Thivolle-Cazat, J., Basset, J. M. & Corker, J. Synthesis, characterization, and reactivity, in the C–H bond activation of cycloalkanes, of a silica-supported tantalum(III) monohydride complex: $(=SiO)_2Ta^{III}-H$. *J. Am. Chem. Soc.* **118**, 4595–4602 (1996).
27. Nicolai, G. P. & Basset, J. M. Primary selectivity in the activation of the carbon-hydrogen bonds of propane by silica supported zirconium hydride. *Appl. Catal. A* **146**, 145–156 (1996).
28. Sherry, A. E. & Wayland, B. B. Metalloradical activation of methane. *J. Am. Chem. Soc.* **112**, 1259–1261 (1990).
29. Wayland, B. B., Ba, S. & Sherry, A. E. Activation of methane and toluene by rhodium(II) porphyrin complexes. *J. Am. Chem. Soc.* **113**, 5305–5311 (1991).
30. Cummins, C. C., Baxter, S. M. & Wolczanski, P. T. Methane and benzene activation via transient $(^tBu_3SiNH)_2Zr = NSi^tBu_3$. *J. Am. Chem. Soc.* **110**, 8731–8733 (1988).
31. Walsh, P. J., Hollander, F. J. & Bergman, R. G. Generation, alkyne cycloaddition, arene C–H activation, N–H activation, and dative ligand trapping reactions of the first monomeric imidozirconocene $(Cp_2Zr = NR)$ complexes. *J. Am. Chem. Soc.* **110**, 8729–8731 (1988).
32. Bennett, J. L. & Wolczanski, P. T. Selectivities in hydrocarbon activation: kinetic and thermodynamic investigations of reversible 1,2-RH-elimination from $(silox)_2(^tBu_3SiNH)TiR$ ($silox = ^tBu_3SiO$). *J. Am. Chem. Soc.* **119**, 10696–10719 (1997).
33. Schafer, D. F. & Wolczanski, P. T. d^0 alkane complexes $(^tBu_3SiN =)_3W(RH)$ precede C–H activation and formation of $(^tBu_3SiN =)_2(^tBu_3SiNH)WR/R'$. *J. Am. Chem. Soc.* **120**, 4881–4882 (1998).
34. Tran, E. & Legzdins, P. Thermal alkane C–H bond activation by a tungsten alkylidene complex: the reversal of α -hydrogen elimination. *J. Am. Chem. Soc.* **119**, 5071–5072 (1997).
35. Stahl, S. S., Labinger, J. A. & Bercaw, J. E. Homogeneous oxidation of alkanes by electrophilic late transition metals. *Angew. Chem. Int. Edn Engl.* **37**, 2180–2192 (1998).
36. Hall, C. & Perutz, R. Transition metal alkane complexes. *Chem. Rev.* **96**, 3125–3146 (1996).
37. Gefakis, S. & Ball, G. E. Direct observation of a transition metal alkane complex, $CpRe(CO)_2(cyclopentane)$, using NMR spectroscopy. *J. Am. Chem. Soc.* **120**, 9953–9954 (1998).
38. Weiller, B. H., Wasserman, E. P., Bergman, R. G., Moore, C. B. & Pimentel, G. C. Time-resolved IR spectroscopy in liquid rare gases: direct rate measurement of an intermolecular alkane C–H oxidative addition reaction. *J. Am. Chem. Soc.* **111**, 8288–8290 (1989).
39. Bengali, A. A., Schultz, R. H., Moore, C. B. & Bergman, R. G. Activation of the C–H bonds in neopentane and neopentane- d_{12} by $(\eta^5-C_5(CH_3)_5)Rh(CO)_2$: spectroscopic and temporal resolution of rhodium–krypton and rhodium–alkane complex intermediates. *J. Am. Chem. Soc.* **116**, 9585–9589 (1994).
40. Gross, C. L. & Girolami, G. S. Metal-alkane complexes. Rapid exchange of hydrogen atoms between hydride and methyl ligands in $[(C_5Me_5)Os(dmpm)(CH_3)H]^+$. *J. Am. Chem. Soc.* **120**, 6605–6606 (1998).
41. Heinekey, D. M. & Oldham, W. Coordination chemistry of dihydrogen. *Chem. Rev.* **93**, 913–926 (1993).
42. Bromberg, S. E. *et al.* The mechanism of a C–H bond activation reaction in room-temperature alkane solution. *Science* **278**, 260–263 (1997).
43. Thompson, M. E. *et al.* σ -Bond metathesis for carbon-hydrogen bonds of hydrocarbons and Sc–R (R = H, alkyl, aryl) bonds of permethylcandocene derivatives. Evidence for noninvolvement of the π system in electrophilic activation of aromatic and vinylic C–H bonds. *J. Am. Chem. Soc.* **109**, 203–219 (1987).
44. Jones, W. D. & Feher, F. J. The mechanism and thermodynamics of alkane and arene carbon-hydrogen bond activation in $(C_5Me_5)Rb(PMe_3)(R)H$. *J. Am. Chem. Soc.* **106**, 1650–1663 (1984).
45. Janowicz, A. H. & Bergman, R. G. Activation of C–H bonds in saturated hydrocarbons on photolysis of $(\eta^5-C_5Me_5)(PMe_3)IrH_2$: relative rates of reaction of the intermediate with different types of C–H bonds and functionalization of the metal-bound alkyl-groups. *J. Am. Chem. Soc.* **105**, 3929–3939 (1983).
46. McNamara, B. K., Yeston, J. S., Bergman, R. G. & Moore, C. B. The effect of alkane structure on rates of photoinduced C–H bond activation by $Cp^*Rh(CO)_2$ in liquid rare gas media: An infrared flash kinetics study. *J. Am. Chem. Soc.* **121**, 6437–6443 (1999).
47. Shilov, A. E. & Shulpin, G. B. Activation and catalytic reactions of alkanes in solutions of metal complexes. *Russ. Chem. Rev.* **56**, 442–464 (1987).
48. Luinstra, G. A., Wang, L., Stahl, S. S., Labinger, J. A. & Bercaw, J. E. C–H activation by aqueous platinum complexes: a mechanistic study. *J. Organometall. Chem.* **504**, 75–91 (1995).
49. Shilov, A. E. *Activation of Saturated Hydrocarbons by Transition Metal Complexes* (Reidel, Boston, 1984).
50. Labinger, J. A., Herring, A. M., Lyon, D. K., Luinstra, G. A., Bercaw, J. E., Horváth, I. & Eller, K. Oxidation of hydrocarbons by aqueous platinum salts: mechanism and selectivity. *Organometallics* **12**, 895–905 (1993).
51. Sen, A., Benvenuto, M. A., Lin, M., Hutson, A. C. & Basickes, N. Activation of methane and ethane and their selective oxidation to the alcohols in protic media. *J. Am. Chem. Soc.* **116**, 998–1003 (1994).
52. Freund, M. S., Labinger, J. A., Lewis, N. S. & Bercaw, J. E. Electrocatalytic functionalization of alkanes using aqueous platinum salts. *J. Mol. Catal.* **87**, L11–L16 (1994).
53. Geletii, Y. V. & Shilov, A. E. Catalytic oxidation of alkanes by molecular oxidation. Oxidation of methane in the presence of platinum salts and heteropoly acids. *Kinetics Catal. Engl. Trans.* **24**, 413–416 (1983).
54. Lin, M., Shen, C., Garcia-Zayas, E. A. & Sen, A. Catalytic Shilov chemistry: Platinum chloride-catalyzed oxidation of terminal methyl groups by dioxygen. *J. Am. Chem. Soc.* **123**, 1000–1001 (2001).
55. Holtcamp, M. W., Henling, L. M., Day, M. W., Labinger, J. A. & Bercaw, J. E. Intramolecular and intermolecular C–H activation at a cationic platinum(II) center. *Inorg. Chim. Acta* **270**, 467–478 (1998).
56. Johansson, L., Tilst, M., Labinger, J. A. & Bercaw, J. E. Mechanistic investigation of benzene C–H activation at a cationic platinum(II) center. Direct observation of a platinum(II) benzene adduct. *J. Am. Chem. Soc.* **122**, 10846–10855 (1997).
57. Rostovtsev, V. V., Labinger, J. A., Bercaw, J. E., Lasseter, T. L. & Goldberg, K. I. Oxidation of dimethylplatinum(II) complexes with dioxygen. *Organometallics* **17**, 4530–4531 (1998).
58. Periana, R. A. *et al.* Platinum catalysts for the high-yield oxidation of methane to a methanol derivative. *Science* **280**, 560–564 (1998).
59. Waltz, K. M. & Hartwig, J. F. Selective functionalization of alkanes by transition-metal boryl complexes. *Science* **277**, 211–231 (1997).
60. Chen, H. Y. & Hartwig, J. F. Catalytic, regiospecific end-functionalization of alkanes: rhenium-catalyzed borylation under photochemical conditions. *Angew. Chem. Int. Edn Engl.* **38**, 3391–3393 (1999).
61. Chen, H., Schlecht, S., Semple, T. C. & Hartwig, J. F. Thermal, catalytic, regiospecific functionalization of alkanes. *Science* **287**, 1995–1997 (2000).
62. Iverson, C. N. & Smith, M. R. Stoichiometric and catalytic B–C bond formation from unactivated hydrocarbons and boranes. *J. Am. Chem. Soc.* **121**, 7096–7097 (1999).
63. Crabtree, R. H., Mihelcic, J. M. & Quirk, J. M. Iridium complexes in alkane dehydrogenation. *J. Am. Chem. Soc.* **101**, 7738–7739 (1979).
64. Jensen, C. M. Iridium PCP pincer complexes: highly active and robust catalysts for novel homogeneous aliphatic dehydrogenations. *Chem. Commun.* 2443–2449 (1999).
65. Liu, F. C., Pak, E. B., Singh, B., Jensen, C. M. & Goldman, A. S. Dehydrogenation of n-alkanes catalyzed by iridium "pincer" complexes: regiospecific formation of alpha-olefins. *J. Am. Chem. Soc.* **121**, 4086–4087 (1999).
66. Liu, F. C. & Goldman, A. S. Efficient thermochemical alkane dehydrogenation and isomerization catalyzed by an iridium pincer complex. *Chem. Commun.* 655–656 (1999).

Correspondence and requests for materials should be addressed to J.A.L. (e-mail: jal@its.caltech.edu).

Crystal structure and mechanism of a calcium-gated potassium channel

Youxing Jiang, Alice Lee, Jiayun Chen, Martine Cadene, Brian T. Chait & Roderick MacKinnon

Howard Hughes Medical Institute, Laboratory of Molecular Neurobiology and Biophysics and Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

Ion channels exhibit two essential biophysical properties; that is, selective ion conduction, and the ability to gate-open in response to an appropriate stimulus. Two general categories of ion channel gating are defined by the initiating stimulus: ligand binding (neurotransmitter- or second-messenger-gated channels) or membrane voltage (voltage-gated channels). Here we present the structural basis of ligand gating in a K^+ channel that opens in response to intracellular Ca^{2+} . We have cloned, expressed, analysed electrical properties, and determined the crystal structure of a K^+ channel (MthK) from *Methanobacterium thermoautotrophicum* in the Ca^{2+} -bound, opened state. Eight RCK domains (regulators of K^+ conductance) form a gating ring at the intracellular membrane surface. The gating ring uses the free energy of Ca^{2+} binding in a simple manner to perform mechanical work to open the pore.

Ion channels are central to a wide range of biological processes including cell volume regulation, movement and electrical signal generation¹. Ion channel proteins span the membrane of a cell, forming a conduction pathway, or pore, through which ions diffuse down their electrochemical gradient across the membrane. To understand how an ion channel operates as a molecular machine we addressed two mechanistic issues. First, how do ions flow selectively through the pore, and second, how does the pore gate, or open, in response to the appropriate stimulus? For K^+ ion channels, significant progress has been made towards understanding the mechanism of selective ion conduction^{2–4}. Here we address the mechanism of opening in a ligand-gated K^+ channel.

K^+ channels belong to a family of ion channels called tetrameric cation channels. The family includes K^+ , Na^+ , Ca^{2+} , cyclic nucleotide-gated, and several other ion channels. They contain four membrane-spanning subunits or domains surrounding a central pore that is selective for cations of one kind or another. On the basis of the KcsA K^+ channel structure^{2,4}, it seems that cation selectivity is an intrinsic property of the pore architecture, which provides a special arrangement of cation-attractive 'pore' α -helices probably shared by all tetrameric cation channel family members.

Gating in the tetrameric cation channels is conferred through the attachment of gating domains to the pore. In channels whose gate opens in response to the membrane voltage (voltage-dependent channels), an integral membrane 'voltage sensor' domain is present on each subunit^{5,6}. In ligand-gated channels, ligand-binding domains are attached to the pore in the aqueous solution near the membrane surface^{7–11} (Fig. 1). The basic function of these gating domains is to perform mechanical work on the ion conduction pore to change its conformation between closed and opened states. Thus, a voltage sensor converts energy stored in the membrane electric field into mechanical work, whereas ligand-binding domains convert the free energy of ligand binding into mechanical work. Thus, ion channel gating comes down to electromechanical or chemomechanical coupling between a gating unit and the pore unit.

Regulators of K^+ conductance (RCK) domains are found in many ligand-gated K^+ channels, most often attached to the intracellular carboxy terminus (Fig. 2a)¹¹. These domains are prevalent among prokaryotic K^+ channels, and are also found in eukaryotic, high-conductance Ca^{2+} -activated K^+ channels (BK channels). Several RCK domain amino-acid sequences are shown in Fig. 2a. The X-ray structure of one RCK domain, that from an *Escherichia coli* K^+ channel, reveals an α - β -protein with a fold similar to dehydrogenase

enzymes¹¹. The domain forms a homodimer, producing a cleft between two lobes. The dimer is similar in its structure to certain bilobed, amino acid and nutrient molecule-binding proteins found in the periplasm of bacteria^{12–14}. We postulated that the *E. coli* K^+ channel binds a ligand molecule in the cleft to affect channel gating. On the basis of amino-acid sequence analysis it is apparent that certain RCK domains bind nicotinamide adenine dinucleotide (NAD)^{15,16}, but for many RCK domains the ligand is unknown.

Analysis of DNA sequences shows that RCK domains occur in at least five different contexts (Fig. 2b): (1) as a single domain on the C terminus of a K^+ channel (for example, many prokaryotic K^+ channels); (2) as two tandem RCK domains on the C terminus (for example, eukaryotic BK channels); (3) as two domains, one on the N and another on the C terminus; (4) as a soluble protein (not part of a K^+ channel gene) consisting of two tandem RCK domains; and (5) as a soluble protein consisting of a single RCK domain. Of note, in three out of five contexts (Fig. 2b, (2)–(4)), RCK domains occur in pairs.

The MthK channel

We cloned and expressed a K^+ channel gene from the archaeon *Methanobacterium thermoautotrophicum*. The gene codes for a K^+ channel that we have named MthK; it contains two membrane-spanning segments per subunit, which forms one subunit of the transmembrane pore, and a single C-terminal RCK domain (Fig. 3a). Expression in *E. coli*, extraction with decylmaltoside

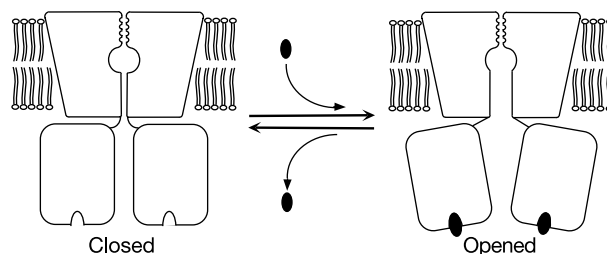


Figure 1 Ligand-activated ion channel gating. A ligand (black oval) binds to receptor domains, inducing a conformational change that leads to opening of the ion conduction pore. The ligand receptor is usually located at the membrane surface on the extracellular side (top) in neurotransmitter-gated ion channels and on the intracellular side (bottom) in second-messenger-gated ion channels, as shown here.

(DM), and purification yields a protein–detergent complex that elutes in a peak at approximately 11 ml on a superdex-200 (10/30) gel filtration column (Fig. 3b). Protein from this peak runs in two bands near a relative molecular mass of 26,000 and 200,000 (M_r 26K and 200K, respectively) on SDS–polyacrylamide gel electrophoresis (PAGE) (Fig. 3b). Mass spectrometry shows that the approximately 200K band contains the full-length gene product, presumably migrating as a tetramer on SDS–PAGE, and that the roughly 26K band contains only the C-terminal RCK domain, beginning at amino-acid position Met 107. In separate experiments we observed a similar result on overexpression of the *E. coli* K^+ channel in *E. coli*: two proteins were produced, one full length and another corresponding to a C-terminal RCK domain, beginning at an internal methionine residue. We reasoned that a single gene gives rise to two gene products, and that both products are required in the assembly of a functional K^+ channel (Fig. 3d). Mutation of the hypothesized internal start site, Met 107, to Ile (M107I) eliminated the 26K gene product (Fig. 3c) and resulted in reduced levels of MthK channel expression. Curiously, on the superdex-200 column the mutant channel elutes at approximately 10 ml, indicating that it is larger in size than the wild-type channel (Fig. 3c). The structural analysis below offers an explanation for this unexpected result; that two mutant channels bind to each other once they have been extracted from the membrane (Fig. 3e). Such binding would provide compensation for the missing soluble domains and explain the large protein–detergent complex on gel filtration.

Electrophysiological analysis

MthK channels produced with Met 107 intact were reconstituted into planar lipid bilayers of 1-palmitoyl-2-oleoyl phosphatidylglycerol (POPG) and 1-palmitoyl-2-oleoyl phosphatidylethanolamine

(POPE) (Fig. 4). The channel is selective for K^+ over Na^+ (data not shown), and in solutions containing 150 mM KCl and 10 mM HEPES, pH 7.0 on both sides of the membrane it exhibits inward rectification: the current–voltage curve is nonlinear, with a greater slope at negative voltages (Fig. 4a). Single channels show a conductance of approximately 200 pS at -100 mV. The K^+ channel blocker charybdotoxin (CTX), a protein from scorpion venom¹⁷, inhibits the MthK channel, underscoring the structural similarity between this prokaryotic K^+ channel and its eukaryotic family members (Fig. 4b). CTX binds to the extracellular pore entryway of K^+ channels¹⁸, which allows easy assessment of a channel's orientation within the membrane, that is, it distinguishes the extracellular (toxin sensitive) and intracellular (toxin insensitive) sides.

Figure 4b shows an important property of the MthK channel. When Ca^{2+} is applied to the intracellular solution, the probability of channel opening increases (Fig. 4b). The '0 mM' Ca^{2+} recording contains trace amounts of Ca^{2+} necessary to open channels; in the presence of EDTA, a Ca^{2+} chelator, MthK channels essentially never open. Thus, Ca^{2+} regulates or gates the channel from the intracellular solution: the effect of Ca^{2+} continues into the millimolar concentration range. The reduced single-channel amplitude at higher Ca^{2+} concentrations is due to rapid pore block by Ca^{2+} ions, a well known phenomenon^{19–21}. The principal result, however, is that Ca^{2+} gates the channel open. We do not know whether Ca^{2+} is a physiological ligand for gating the MthK channel, but for the purposes of the present study, Ca^{2+} is a ligand that opens the channel in a concentration-dependent manner.

X-ray structure analysis

We solved the three-dimensional structure of the MthK channel by X-ray crystallography at a resolution of 3.3 Å (Methods; see also

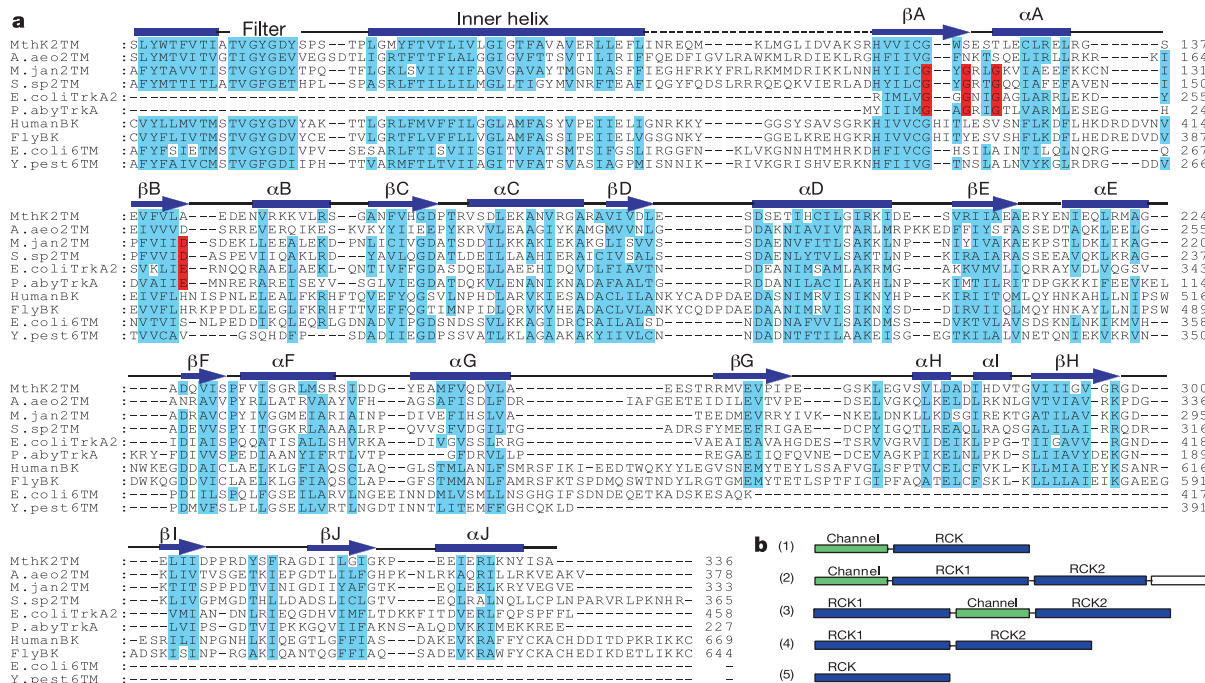


Figure 2 Sequence analysis of proteins containing RCK domains. **a**, Partial sequence alignment of K^+ channels and prokaryotic K^+ transporters (TrkA). The alignment begins at the K^+ channel signature sequence (Filter) and ends at the MthK channel (MthK2TM) C terminus. Secondary structure assignment is based on the MthK crystal structure: blue bars and arrows show α -helices and β -strands, respectively; dashed line shows the structurally undefined linker connecting the pore to the RCK domain. Cyan indicates semi-conserved sequence; red shows an NAD-binding motif present in some RCK domains. K^+ channels: MthK2TM, *M. thermotrophicum* (Gl:2622639); A.aeo2TM, *Aquifex aeolicus*

(Gl:2983007); M.jan2TM, *Methanococcus jannaschii* (Gl:1498918); S.sp2TM, *S. sp* (Gl:7447545); E.coli6TM, *E. coli* (Gl 400124); Y.pest6TM, *Yersinia pestis* (sanger_632); TrkA proteins: E.coliTrkA2, *E. coli*, domain 2 (Gl:136235); P.abbyTrkA, *Pyrococcus abyssi* (Gl:7450648); BK K^+ channels: HumanBK, *Homo sapiens* (Gl:2570854); FlyBK, *Drosophila melanogaster* (Gl:7301192). **b**, RCK domains: (1) single C-terminal domain of K^+ channels; (2) tandem C-terminal domains (BK channels); (3) N- and C-terminal domains (*Synechocystis* sp. channel); and (4) and (5) soluble containing tandem (*E. coli* TrkA) or single (*P. abyssi* TrkA) domains of prokaryotic K^+ transporters.

Supplementary Information Table 1). Crystals were obtained using the mutant M107I in which the proposed second start site was removed (Fig. 3a, c). Within a P6₂22 crystal lattice, two channels are arranged in a 'back-to-back' fashion (Fig. 3e). From our biochemical studies (Fig. 3) and subsequent structural analysis we conclude that the back-to-back channels in the crystal share RCK domains, and that a functional K⁺ channel corresponds to the diagram in Fig. 3d. Thus, in the membrane an MthK channel apparently has eight RCK domains at its intracellular surface, four co-assembling from solution. We refer to the membrane-spanning portion of the MthK channel as the pore, and the structure formed by the eight RCK domains as the gating ring.

The electron density map in the region of the gating ring is of high quality, enabling us to build nearly all amino-acid side chains. The pore is not as well ordered in the crystal, but four-fold noncrystallographic symmetry (NCS) averaging provided good definition of the main chain helical structure. The pore was built as a polyaniline model, but density for the largest side chains and sequence analysis permitted the unambiguous assignment of the helical register within the pore. The final MthK channel model (R_{cryst} 29.9%; R_{free} 31.5%) is built from amino acids 19–336, with the exception of 17 unresolved residues (99–115) at the junction between the pore and the gating ring (Fig. 5a).

A surface-rendered structure of the MthK channel with one subunit removed is shown with the membrane-spanning pore above and the intracellular gating ring below (Fig. 5b). The entire channel, pore and gating ring, shares four-fold symmetry by rotation of the unit shown in Fig. 5a about the pore axis. The gating ring is large (approximately 1,800 amino acids) compared with the pore (about 400 amino acids), and the hole down the centre of the gating ring is approximately 20 Å in diameter, providing a wide path for

the movement of K⁺ ions between the cytoplasm and the pore.

The crystals were grown in the presence of 200 mM Ca²⁺, a ligand that opens the channel in electrophysiological experiments (Fig. 4b). Well defined peaks in the electron density map are most consistent with Ca²⁺ ions on the basis of local chemistry; these ions are shown as yellow spheres lining the wall inside the gating ring (Fig. 5b). Two Ca²⁺ ions are located at the base of a cleft between two RCK domains (Fig. 5a), coordinated by the carboxylate groups of glutamate residues 210 and 212, and by aspartate residue 184 (Fig. 5c). To test whether these Ca²⁺ ions in the structure are directly correlated with channel gating, we mutated Asp 184 to Asn in the Ca²⁺ binding site, and reconstituted the mutant channel (Met 107 intact) into lipid bilayers. Channel gating was affected signifi-

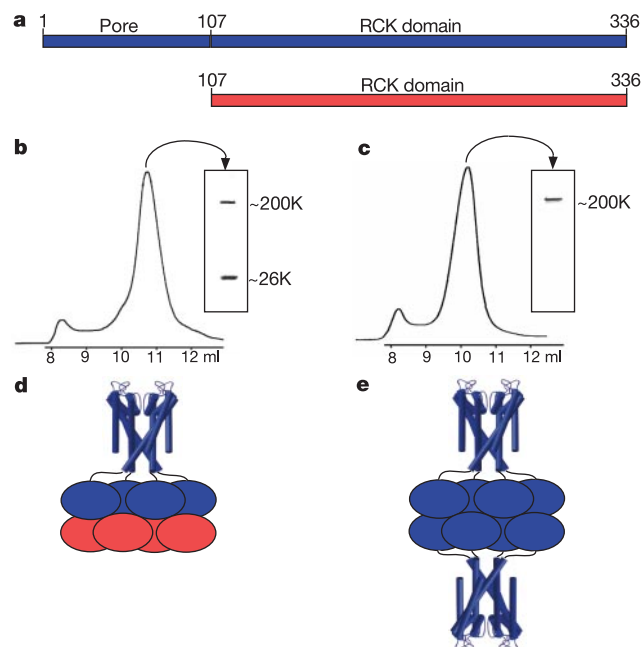


Figure 3 Biochemical analysis of MthK. **a**, Expression of wild-type MthK yields two proteins: full-length channel (blue bar) and the RCK domain from Met 107 (red bar). **b**, Gel-filtration profile and SDS-PAGE of wild-type MthK purified in DM. Major peak elutes at 10.7 ml (superdex-200, 10/30); protein from this peak migrates as two bands on SDS-PAGE (inset) at 26K (RCK domain) and about 200K (MthK tetramer). **c**, Gel-filtration profile and SDS-PAGE of MthK mutant M107I purified in DM. Major peak elutes at 10.2 ml; protein from this peak migrates as a single band on SDS-PAGE (inset) at about 200K (MthK tetramer). **d, e**, Hypothesis explaining the biochemical results for wild-type (**d**) and M107I mutant (**e**) channel expression. Pore is shown as cylinders and RCK domains as ovals with colours as in **a**.

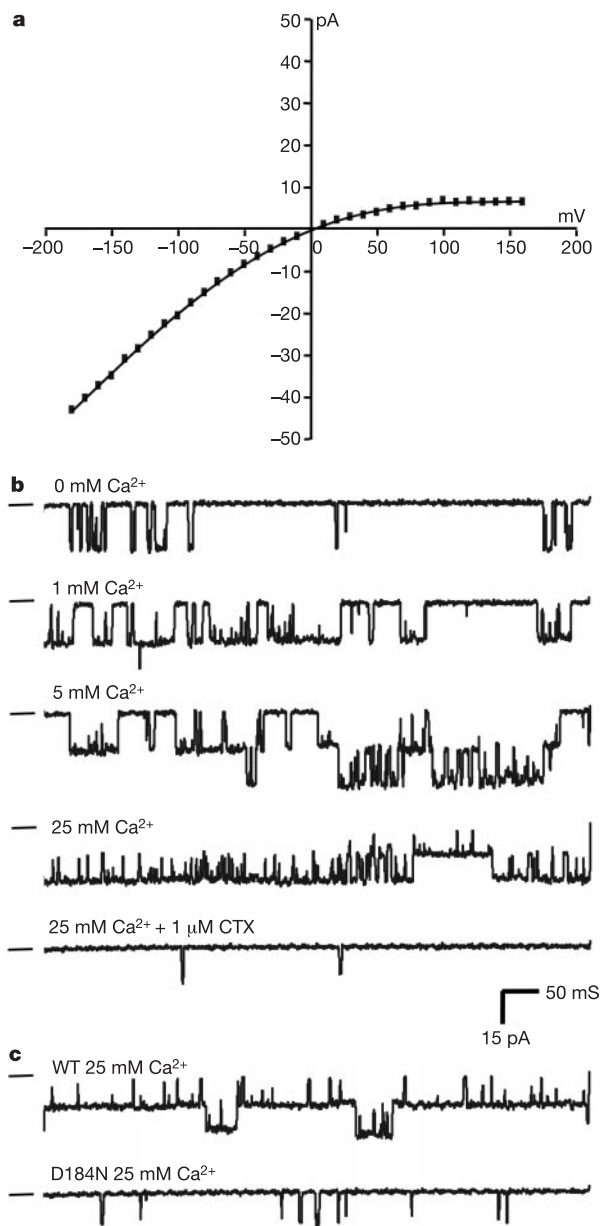


Figure 4 Electrophysiological analysis of MthK in lipid bilayers. **a**, Current–voltage curve of MthK measured in solutions of 150 mM KCl, 10 mM HEPES, pH 7.0, on both sides of the membrane. **b**, Single channel traces of MthK in response to various internal Ca²⁺ concentrations or external CTX. Membrane voltage was –100 mV so that channel openings cause a downward current; lines to the left mark the zero current level. **c**, Single channel traces of wild-type (WT) MthK and D184N mutant in the presence of 25 mM internal Ca²⁺, 150 mM KCl, 10 mM HEPES, pH 7.0.

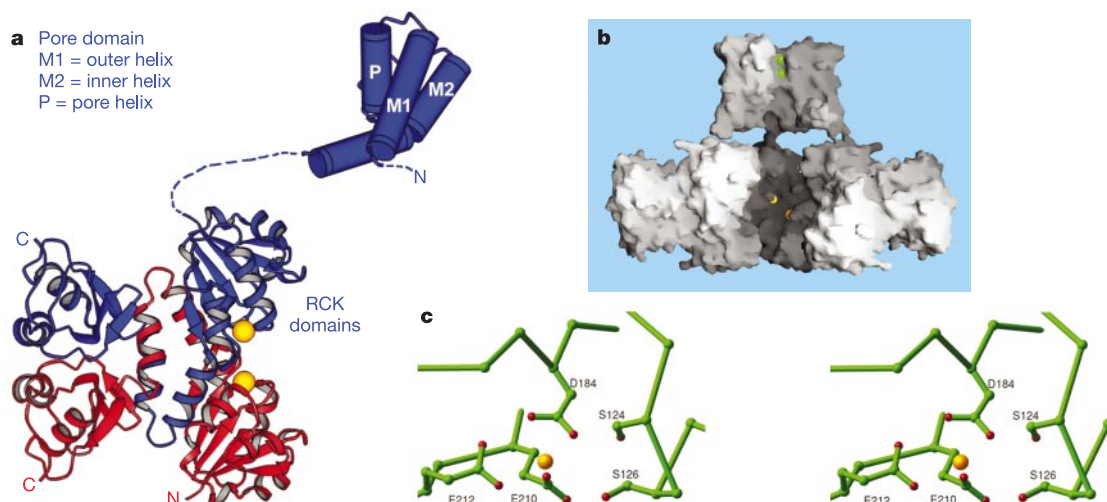


Figure 5 Structure of the MthK K^+ channel. **a**, A single four-fold repeated unit of the MthK channel contains a pore domain (blue cylinders) and two RCK domains (blue and red ribbons). Disordered segments (dashed) and Ca^{2+} ions (yellow spheres) are shown. **b**, Molecular surface of the MthK tetramer viewed from the side (extracellular surface on

top) with one unit removed. Green and yellow spheres show K^+ and Ca^{2+} ions, respectively. **c**, Stereo diagram of the Ca^{2+} -binding site with protein atoms in green and red and Ca^{2+} in yellow. All ribbon diagrams were prepared using RIBBONS³⁹ and all surface models using GRASP⁴⁰.

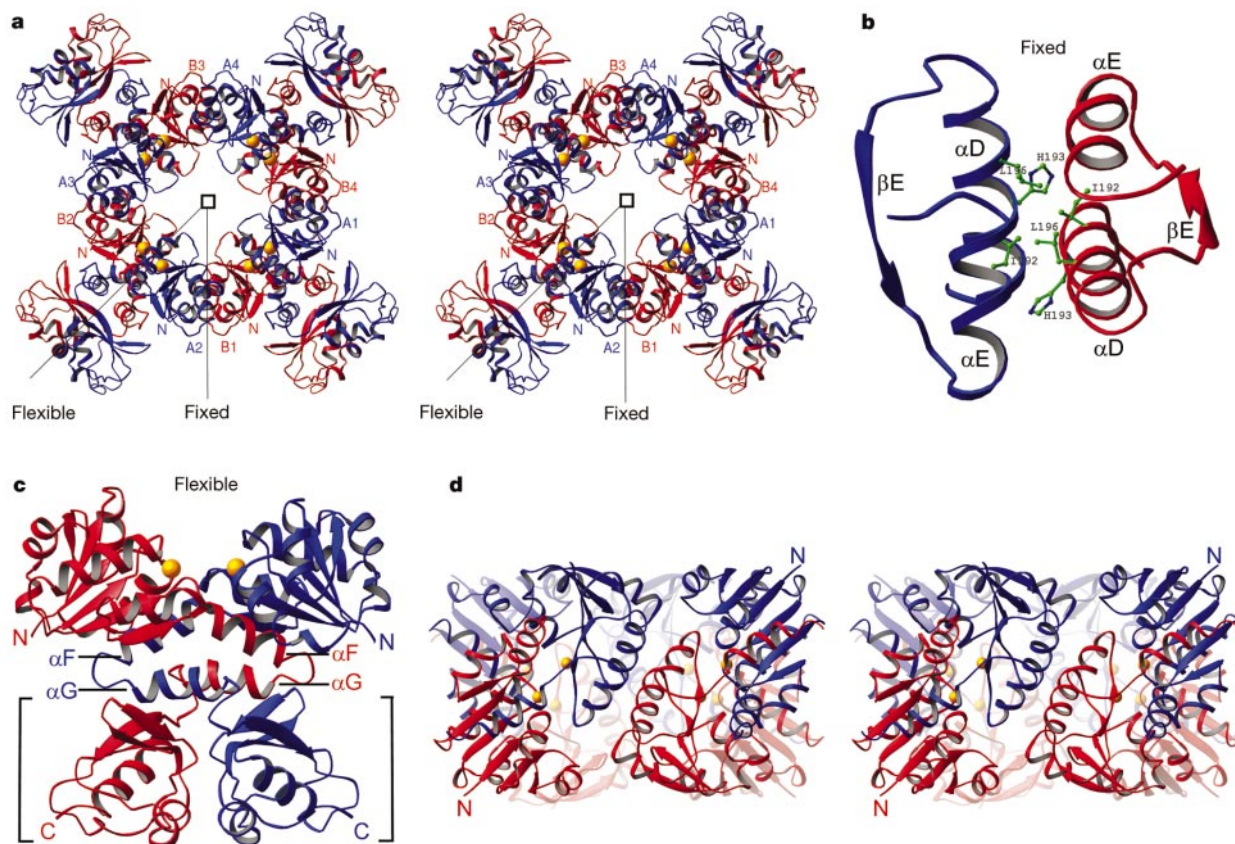


Figure 6 Structure of the gating ring. **a**, Stereo diagram of the gating ring viewed down the four-fold axis (square). Eight RCK domains are divided into two groups, A1–A4 (blue) and B1–B4 (red), with N termini labelled (N). Viewed from the membrane, blue domains would be attached to the pore via a continuous polypeptide chain and red domains assembled from solution. Fixed and flexible dimer interfaces hold the ring together. **b**, Helices αD and αE form the fixed interface. Several amino acids conserved as

hydrophobic or aromatic (Ile 192, His 193 and Leu 196) are shown. **c**, C-terminal subdomains (brackets) and helices αF and αG form the flexible interface. A cleft between two RCK domains creates a ligand-binding site with Ca^{2+} bound at the base (yellow spheres). **d**, Stereo diagram of the gating ring viewed from the side after removing the subdomains.

cantly by this mutation (Fig. 4c). In particular, at a high Ca^{2+} concentration the mean open time was short and the mean closed time long compared with the wild-type channel, resulting in an apparent decrease in the ability of Ca^{2+} to open the channel. We conclude that by binding to its sites in the cleft between RCK domains in the gating ring, Ca^{2+} causes the MthK channel to open. Our crystal structure shows the gating ring in its Ca^{2+} -bound conformation that would favour pore opening. Not surprisingly, the pore of the MthK channel in the crystal is wide open: the arrangement of inner helices is very different from that observed in the KcsA K^+ channel²².

The gating ring

The gating ring has an arrangement of eight identical RCK domains (Fig. 6a). If one imagines that the blue domains—nearest the viewer in Fig. 6a—are attached to the membrane pore through the polypeptide chain, then the red domains come from solution. The complementary interlocking of blue and red domains to form the closed ring shows why a functional MthK channel apparently requires eight RCK domains: blue domains never directly contact blue domains; similarly, red never contact red.

Two unique protein interfaces between blue and red domains hold the ring together. We have called one of these the fixed interface. The fixed interface is formed by helices αD and αE (Fig. 6a, b). Based on the conservation of specific hydrophobic amino acids on helix αD , we think that the fixed interface occurs in nearly all RCK domains. We call the other protein interface flexible (Fig. 6a, c). The flexible interface is formed in part by the subdomains that protrude out from the gating ring, and in part by helices αG and αF . The peripheral subdomain is variable among RCK domains, and appears to be a unit that adds stability to the flexible interface (Fig. 6c, brackets). The crucial and conserved (among RCK domains) aspect of the flexible interface is not the peripheral subdomain, but the region where the ligand-binding cleft is formed between main lobes of two adjacent RCK domains (Fig. 6c, top). This interface forms the Ca^{2+} -binding cleft in the MthK channel-gating ring.

It is instructive to view the gating ring from the side after

removing the peripheral subdomains (Fig. 6d). From this perspective, the alternating arrangement of RCK domains is visible. Going around the ring, left to right on the front face, a blue domain is attached to a red domain through a fixed interface, the red domain is then attached to a blue domain by a flexible interface. This pattern of fixed and flexible interfaces repeats itself around the ring.

A mechanism for Ca^{2+} -activation

How might the gating ring convert the free energy of Ca^{2+} binding into mechanical work on the pore? To address this question, we compared RCK domain structures (without their peripheral subdomains) from the MthK and *E. coli* K^+ channels (Fig. 7). The RCK domain from the MthK channel is from the complete channel structure, solved with Ca^{2+} ions in the ligand-binding site, whereas the *E. coli* channel RCK domain was solved separately from the channel and is without a ligand¹¹. Despite these differences, the same two domain–domain interfaces, fixed and flexible, occur in both crystals, but with notable variation. Across the fixed interface, two RCK domains of the MthK channel (Fig. 7a, black) have the same relative relationship to each other as two *E. coli* channel RCK domains (Fig. 7a, red). That is, two RCK domains across this interface are fixed with respect to each other. Across the flexible interface, however, there are large structural differences: the ligand-binding cleft is relatively closed between two MthK channel RCK domains (Fig. 7b, black), and open between two *E. coli* channel RCK domains (Fig. 7b, red). Several reasons can account for the relative movements at the flexible interface: forces imposed by unique crystal lattices, the constraint of being attached to a channel and held in a ring, and the presence of a ligand. The important point is, comparison of the two different crystals indicates that one RCK domain interface is fixed and the other is clearly flexible. Moreover, the flexible hinge, which allows movement, occurs at the base of the cleft near to where Ca^{2+} binds in the MthK channel. One can imagine Ca^{2+} influencing the structure by binding in the cleft. In bacterial periplasmic-binding proteins^{12–14} and glutamate receptor ligand-binding domains^{23,24}, this is exactly what happens; a ligand binds in the cleft between two domains and causes the domains to move relative to each other.

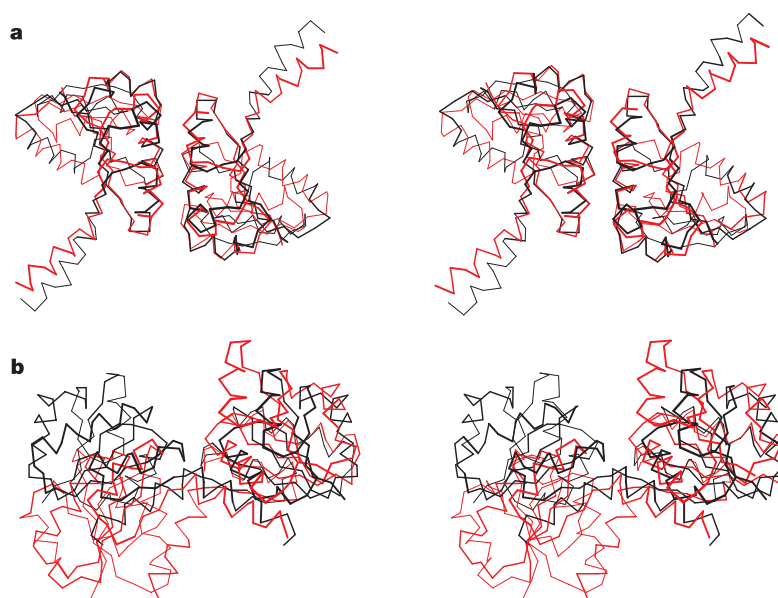


Figure 7 Comparison of different RCK domains reveals possible gating motions. **a**, **b**, RCK domains from the MthK channel (PDB code 1LNQ), residues 116–245 (black C α trace), and the *E. coli* K^+ channel (PDB code 1ID1), residues 244–379 (red C α trace), are superimposed, showing the fixed (**a**) and flexible (**b**) dimer interfaces in stereo. Across the

fixed interface both dimer halves are simultaneously superimposable. Across the flexible interface only one half can be superimposed at once due to relative domain movements. The figures were prepared using Bobscrip⁴¹.

The static compared with dynamic property of the two interfaces offers a plausible explanation for how the gating ring could open and close the channel. In Fig. 8a the crystal structure of the complete MthK channel is shown from the side. The peripheral subdomains have been removed from the gating ring to simplify the picture, and dashed lines show the expected path of the unresolved connection between the pore and gating ring. RCK domains on either side of a fixed interface are the same colour to reinforce the idea that such pairs are stuck rigidly together, and are expected to move as a single rigid unit. Viewed in this way, the gating ring can be described as an assembly of four rigid units (RCK domain pairs attached by the fixed interface), tilted and packed in a left-handed 'bundle'. In principle, the shape of the gating ring is adjustable. To see how, imagine that each rigid unit is rotated about an axis through the fixed interface (coincident with the line labelled 'fixed' in Fig. 6a), and pushed towards the centre of the ring, as shown in Fig. 8b, c; then the entire gating ring will become taller and its diameter reduced (see movie in Supplementary Information). For such rigid unit movements to occur, the flexible interfaces that connect the rigid units around the ring would have to undergo a conformational change. We propose that this is the essence of the machine: movements of the rigid units will reshape the cleft and presumably influence the Ca^{2+} binding affinity, which is energetically equivalent to saying that Ca^{2+} binding will influence movements of the rigid

units. We propose that the gating ring converts the free energy of ligand binding into mechanical work in just this way: Ca^{2+} binds and reshapes the cleft, causing the rigid units to tilt, which in turn expands the diameter of the gating ring and pulls open the pore's inner helices. The manual rigid conformational change shown in Fig. 8b was made without producing large steric clashes and reduces the diameter of the gating ring by about 13 Å. The change in ring diameter is of the correct magnitude to account for the structural differences observed between the inner helices of MthK (an opened channel) and KcsA (a closed channel), as described in ref. 22.

Ligand-induced activation of ion channels often occurs with a high degree of cooperativity²⁵. A structural basis for understanding cooperativity is apparent in the proposed mechanism of the gating ring. When Ca^{2+} binds in one cleft to reshape its structure, the structure of adjacent clefts must also change through movements of the rigid units. Thus, Ca^{2+} binding at one site will have energetic consequences for Ca^{2+} binding to neighbouring sites, giving rise to cooperativity.

The gating ring in MthK channels is assembled from eight RCK domains, one from each of the four pore-forming subunits and four from the intracellular solution. One can imagine in some cases that the four RCK domains from solution might arise from a separate gene coding for an RCK domain. Indeed, there are many prokaryotic genes coding for single, soluble RCK domains. Perhaps some of

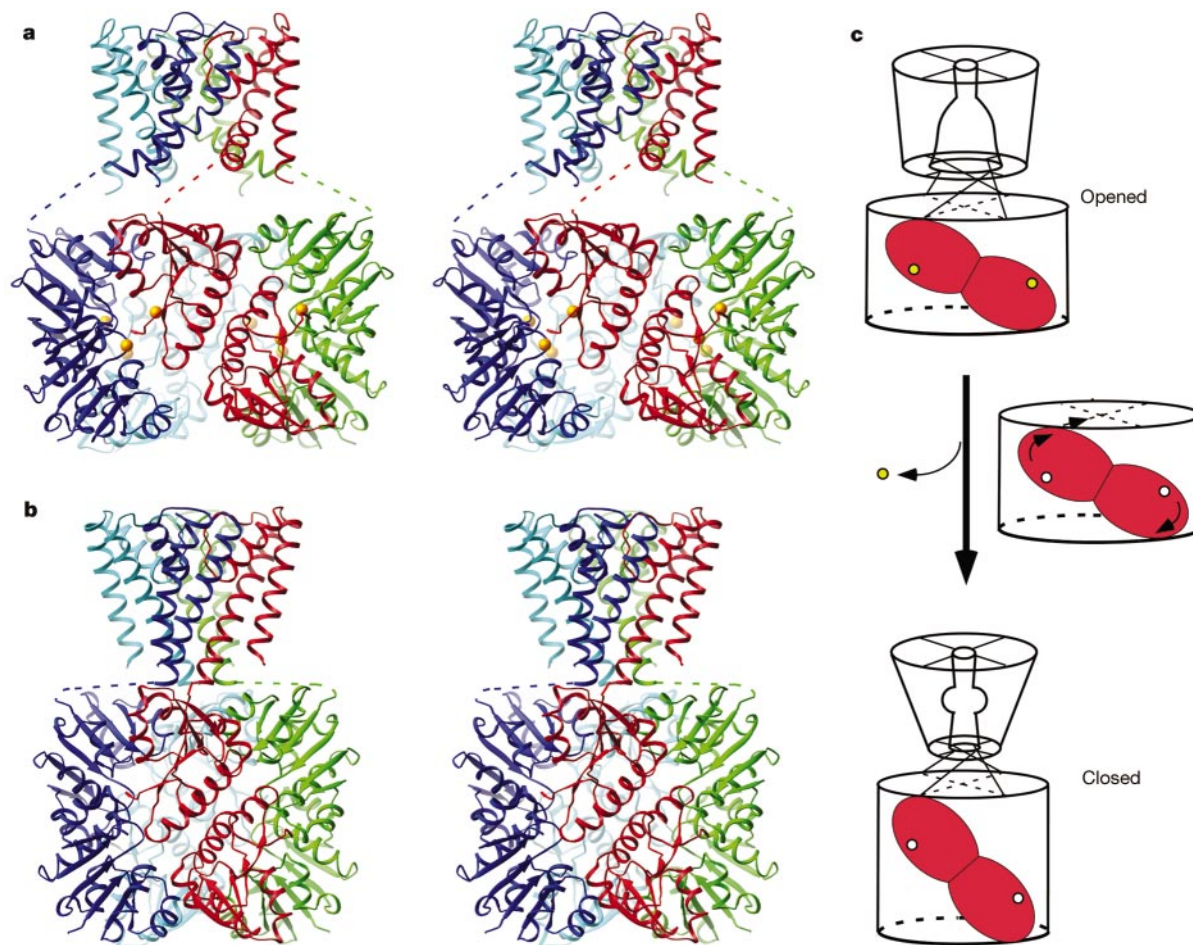


Figure 8 Proposed mechanism of gating. **a**, Crystal structure of the MthK channel (opened conformation) with subdomains removed from the gating ring. RCK domains forming a fixed interface with each other are the same colour. Disordered linkers are shown as dashed lines and Ca^{2+} ions as yellow spheres. **b**, Hypothetical model of the MthK channel in its closed conformation. RCK domain pairs of the same colour were moved as rigid bodies. The movements decrease the gating ring diameter by 13 Å,

allowing the inner helices to achieve the conformation observed in the KcsA K^{+} channel (a closed K^{+} channel). **c**, Cartoon outlining the gating motions. A single rigid unit is shown (red). Arrows show proposed direction of rotation and translation. Release of Ca^{2+} ions (yellow circles) is associated with conformational changes in the gating ring that close the channel.

these domains assemble to form heteromultimer channels with unique ligand-gating properties. Yet another variation on the gating ring apparently occurs in eukaryotic BK channels. These Ca^{2+} -activated K^+ channels have a pair of tandem RCK domains on their C terminus. Therefore, eight RCK domains provided by the four channel subunits probably form a gating ring similar to that in the MthK channel. BK channels have an additional C-terminal domain, following the second RCK domain, which is also important for Ca^{2+} -activation⁹.

Discussion

We have addressed the question of how the binding of a ligand is able to open the pore of an ion channel. We cloned, expressed, studied electrophysiological properties, and determined the crystal structure of a K^+ channel that contains RCK domains, and that is gated by intracellular Ca^{2+} ions. Calcium ions are present in the structure and the channel is in the open conformation. By interacting through two unique protein–protein interfaces, one fixed and the other flexible, eight RCK domains assemble to form a gating ring at the intracellular membrane surface. Through ligand-induced conformational changes at the flexible interface, we propose that the gating ring changes its shape in such a way that it can perform work on the pore, pulling open the inner helices to permit ion conduction. The gating ring is an example of how a seemingly complicated function—the opening of an ion channel through specific ligand binding—is carried out through a simple set of structural rules. □

Methods

Protein preparation and analysis

The gene coding for the MthK K^+ channel was cloned by polymerase chain reaction (PCR) amplification of genomic DNA from *M. thermoautotrophicum* (ATCC). The gene was inserted into the protein expression vector pQE70 (Qiagen) between *SphI* and *BglII* restriction endonuclease sites with a thrombin cleavage site between a C-terminal hexahistidine sequence and the channel. In a mutated version of the gene, site-directed mutagenesis using QuickChange (Stratagene) was used to replace Met 107 with Ile (M107I mutant). For crystallization, the M107I mutant was expressed in *E. coli* SG13009 (pREP4) cell cultures on induction with 0.4 mM isopropyl- β -D-thiogalactopyranoside (IPTG). Expressed protein was extracted from cell lysate using 40 mM decylmaltoside (DM) and purified on a Talon Co^{2+} affinity column (Clontech). The protein was maintained in a solution of 5 mM DM, 20 mM Tris, pH 8.0 and 100 mM KCl. Nonspecifically bound protein was washed using 20 mM imidazole added to the above buffer, and the channel was then eluted with 300 mM imidazole. Immediately after elution, 1.0 unit of thrombin (Roche) per 3.0 mg channel was added to cleave the hexahistidine sequence overnight at room temperature. Protein was concentrated to about 15 mg ml⁻¹, run on a Superdex-200 (10/30) column (Pharmacia) in 5 mM N,N-dimethyldodecylamine-N-oxide (LDAO), 20 mM Tris, pH 8.0 and 100 mM KCl, and concentrated to 15.0 mg ml⁻¹ (Centricon YM-50) for crystallization. For bilayer experiments, the same preparation was carried out using the wild-type channel and a mutant of the wild-type channel (D184N), except in the final gel filtration step where 5.0 mM DM detergent was used. Maldi time-of-flight mass spectrometry^{26,27} (PerSeptive Biosystems Voyager-STR) was used to confirm that the wild-type MthK channel contains two polypeptides corresponding to full length and a partial length from position 107 to 336.

Crystallization

Crystals were grown by sitting-drop vapour diffusion at 20 °C by mixing equal volumes of protein and reservoir which contained 23–26% polyethylene glycol 350 monomethyl ether (PEGMME), 100 mM MES, pH 6.5 and 200 mM CaCl_2 . Crystals were of space group *P6*₃/22, with cell dimensions $a = b = 137 \text{ \AA}$, $c = 373 \text{ \AA}$, $\alpha = \beta = 90^\circ$ and $\gamma = 120^\circ$, and four subunit copies in the asymmetric unit. Cryoprotection was achieved by raising the PEGMME concentration to about 40% by addition to the reservoir, after which crystals were frozen in liquid propane.

Structure determination

Data were collected at -180°C under nitrogen stream. Numerous crystals were studied to obtain the best achievable diffraction data. The native data set used in the structure refinement was obtained at the National Synchrotron Light Source (NSLS) X-25 on a single crystal from protein in which residue 76 was mutated to cysteine; whether the mutation was responsible for modestly improved diffraction is unclear. The initial phases were estimated from CH_3HgCl derivative data collected at the Advanced Light Source (ALS) 5.0.2. All data sets were processed with DENZO and SCALEPACK²⁸. Heavy atom positions for a first derivative were determined by Patterson search using SOLVE²⁹ and by direct methods using SnB³⁰, and for two additional derivatives using difference Fourier methods (CCP4 (ref. 31)). Heavy atom positions were refined and phases calculated using MLPHARE³¹ and SHARP³²; initial phases were improved by solvent flattening with

SOLOMON³³. Model building was aided by identifying the location of RCK domains in the electron density map with FFEEAR³⁴ using the *E. coli* RCK domain (Protein Data Bank code 1ID1, main chain from 241 to 365) as a search model³¹. Noncrystallographic symmetry (NCS) operators for the RCK domains were determined and subsequent four-fold averaging using RAVE³⁵ allowed phase extension from 4.2 Å to 3.3 Å. The NCS operators for the pore were slightly different than for the RCK domains so that the two regions of the electron density map were averaged separately. The model was built using O³⁶ and refinement carried out using CNS³⁷ by iterative cycles of simulated annealing with torsion-angle dynamics and model rebuilding. For convenience, the structure was refined against data reduced to space group *P6*₃, with two complete tetramers in the asymmetric unit. Strong NCS constraints were maintained throughout refinement. Electron density for the RCK domains was high quality, allowing the construction of a complete model from position 116 to 336, including Ca^{2+} ions. The pore was less well defined and built as a polyalanine model from position 19 to 98. The N-terminal 18 residues and residues from 99 to 115 were disordered. The final model contains 16,896 protein atoms, 8 Ca^{2+} ions, 4 K^+ ions, and is refined to R_{cryst} and R_{free} of 29.9% and 31.5%, respectively.

Single channel analysis

Wild-type MthK channels and the mutant D184N were reconstituted from DM into lipid vesicles as described³⁸, to give a final protein concentration of 100 $\mu\text{g ml}^{-1}$. Planar lipid bilayers of POPE (15 mg ml⁻¹) and POPG (5 mg ml⁻¹) in decane were painted over a 300- μm hole in a polystyrene partition separating internal and external solutions. To induce fusion of channel-containing vesicles, solution on the side to which vesicles were added (*cis*) contained 150 mM KCl, 10 mM HEPES, pH 7.0 and the opposite side (*trans*) contained 15 mM KCl, 10 mM HEPES, pH 7.0. After the appearance of channels in the membrane, KCl concentration on the *trans* side was increased to 150 mM. Charybdotoxin was used to define the extracellular side of the channel, and Ca^{2+} concentration was varied on the intracellular side. Membrane voltage was controlled and current recorded using an Axopatch 200B amplifier with a Digidata 1322A analogue-to-digital converter and Axoclamp software (Axon Instrument).

Received 22 March; accepted 22 April 2002.

- Hille, B. *Ion Channels of Excitable Membranes* (Sinauer, Sunderland, Massachusetts, 2001).
- Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K^+ channel–Fab complex at 2.0 Å resolution. *Nature* **414**, 43–48 (2001).
- Morais-Cabral, J. H., Zhou, Y. & MacKinnon, R. Energetic optimization of ion conduction rate by the K^+ selectivity filter. *Nature* **414**, 37–42 (2001).
- Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* **280**, 69–77 (1998).
- Sigworth, F. J. Voltage gating of ion channels. *Q. Rev. Biophys.* **27**, 1–40 (1994).
- Bezanilla, F. The voltage sensor in voltage-dependent ion channels. *Physiol. Rev.* **80**, 555–592 (2000).
- Zagotta, W. N. & Siegelbaum, S. A. Structure and function of cyclic nucleotide-gated channels. *Annu. Rev. Neurosci.* **19**, 235–263 (1996).
- Finn, J. T., Grunwald, M. E. & Yau, K. W. Cyclic nucleotide-gated ion channels: an extended family with diverse functions. *Annu. Rev. Physiol.* **58**, 395–426 (1996).
- Schreiber, M. & Salkoff, L. A novel calcium-sensing domain in the BK channel. *Biophys. J.* **73**, 1355–1363 (1997).
- Xia, X. M. *et al.* Mechanism of calcium gating in small-conductance calcium-activated potassium channels. *Nature* **395**, 503–507 (1998).
- Jiang, Y., Pico, A., Cadene, M., Chait, B. T. & MacKinnon, R. Structure of the RCK domain from the *E. coli* K^+ channel and demonstration of its presence in the human BK channel. *Neuron* **29**, 593–601 (2001).
- Sack, J. S., Saper, M. A. & Quirocho, F. A. Periplasmic binding protein structure and function. Refined X-ray structures of the leucine/isoleucine/valine-binding protein and its complex with leucine. *J. Mol. Biol.* **206**, 171–191 (1989).
- Zou, J. Y., Flocco, M. M. & Mowbray, S. L. The 1.7 Å refined X-ray structure of the periplasmic glucose/galactose receptor from *Salmonella typhimurium*. *J. Mol. Biol.* **233**, 739–752 (1993).
- Bjorkman, A. J. *et al.* Probing protein–protein interactions. The ribose-binding protein in bacterial transport and chemotaxis. *J. Biol. Chem.* **269**, 30206–30211 (1994).
- Bellamacina, C. R. The nicotinamide dinucleotide binding motif: a comparison of nucleotide binding proteins. *FASEB J.* **10**, 1257–1269 (1996).
- Schlösser, A., Hamann, A., Bossemeyer, D., Schneider, E. & Bakker, E. P. NAD⁺ binding to the *Escherichia coli* K(+)-uptake protein TrkA and sequence similarity between TrkA and domains of a family of dehydrogenases suggest a role for NAD⁺ in bacterial transport. *Mol. Microbiol.* **9**, 533–543 (1993).
- Miller, C., Moczydlowski, E., Latorre, R. & Phillips, M. Charybdotoxin, a protein inhibitor of single Ca^{2+} -activated K^+ channels from mammalian skeletal muscle. *Nature* **313**, 316–318 (1985).
- MacKinnon, R. & Miller, C. Mutant potassium channels with altered binding of charybdotoxin, a pore-blocking peptide inhibitor. *Science* **245**, 1382–1385 (1989).
- Moczydlowski, E. & Latorre, R. Gating kinetics of Ca^{2+} -activated K^+ channels from rat muscle incorporated into planar lipid bilayers. Evidence for two voltage-dependent Ca^{2+} binding reactions. *J. Gen. Physiol.* **82**, 511–542 (1983).
- Vergara, C. & Latorre, R. Kinetics of Ca^{2+} -activated K^+ channels from rabbit muscle incorporated into planar bilayers. Evidence for a Ca^{2+} and Ba^{2+} blockade. *J. Gen. Physiol.* **82**, 543–568 (1983).
- Cox, D. H., Cui, J. & Aldrich, R. W. Separation of gating properties from permeation and block in mslo large conductance Ca-activated K^+ channels. *J. Gen. Physiol.* **109**, 633–646 (1997).
- Jiang, Y. *et al.* The open pore conformation of potassium channels. *Nature* **417**, 523–526 (2002).
- Mayer, M. L., Olson, R. & Gouaux, E. Mechanisms for ligand binding to GluR0 ion channels: crystal structures of the glutamate and serine complexes and a closed apo state. *J. Mol. Biol.* **311**, 815–836 (2001).
- Armstrong, N., Sun, Y., Chen, G. Q. & Gouaux, E. Structure of a glutamate-receptor ligand-binding core in complex with kainate. *Nature* **395**, 913–917 (1998).
- Golowasch, J., Kirkwood, A. & Miller, C. Allosteric effects of Mg^{2+} on the gating of Ca^{2+} -activated K^+ channels from mammalian skeletal muscle. *J. Exp. Biol.* **124**, 5–13 (1986).

26. Cadene, M. & Chait, B. T. A robust, detergent-friendly method for mass spectrometric analysis of integral membrane proteins. *Anal. Chem.* **72**, 5655–5658 (2000).
27. Cohen, S. L. & Chait, B. T. Mass spectrometry as a tool for protein crystallography. *Ann. Rev. Biophys. Biomol. Struct.* **30**, 67–85 (2001).
28. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
29. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
30. Weeks, C. M. & Miller, R. The design and implementation of SnB version 2.0. *J. Appl. Cryst.* **32**, 120–124 (1999).
31. Collaborative Computational Project No. 4 The CCP4 suite: programs for X-ray crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
32. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
33. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
34. Cowtan, K. Modified phased translation functions and their application to molecular-fragment location. *Acta Crystallogr. D* **54**, 5, 750–756 (1998).
35. Kleywegt, G. J. & Jones, T. A. in *From the First Map to Final Model* Proceedings of the CCP4 study weekend (eds Bailey, S., Hubbard, R. & Waller, D.) 59–66 (Daresbury Laboratory, Daresbury, 1994).
36. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
37. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
38. Heginbotham, L., LeMasurier, M., Kolmakova-Partensky, L. & Miller, C. Single streptomyces lividans K⁺ channels. Functional asymmetries and sidedness of proton activation. *J. Gen. Physiol.* **114**, 551–560 (1999).
39. Carson, M. Ribbons. *Methods Enzymol.* **277**, 493–505 (1997).
40. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).
41. Esnouf, R. M. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 132–133 (1997).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

We thank the staff at the NSLS, Brookhaven National Laboratory, X-25, Cornell High Energy Synchrotron Source, F1 and ALS, Lawrence Berkeley Laboratory, 5.0.2 for synchrotron support; members of the MacKinnon laboratory, P. Model and M. Russel for discussion; C. Miller and D. Gadsby for critical reading of the manuscript; and W. Chin for help in manuscript preparation. This work was supported by grants from the NIH to R.M. and the National Centre for Research Resources, NIH, to B.T.C. R.M. is an investigator in the Howard Hughes Medical Institute.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.M. (e-mail: mackinn@rockvax.rockefeller.edu). Coordinates have been deposited with the Protein Data Bank under accession code 1LNQ.

The open pore conformation of potassium channels

Youxing Jiang, Alice Lee, Jiayun Chen, Martine Cadene, Brian T. Chait & Roderick MacKinnon

Howard Hughes Medical Institute, Laboratory of Molecular Neurobiology and Biophysics and Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

Living cells regulate the activity of their ion channels through a process known as gating. To open the pore, protein conformational changes must occur within a channel's membrane-spanning ion pathway. KcsA and MthK, closed and opened K^+ channels, respectively, reveal how such gating transitions occur. Pore-lining 'inner' helices contain a 'gating hinge' that bends by approximately 30° . In a straight conformation four inner helices form a bundle, closing the pore near its intracellular surface. In a bent configuration the inner helices splay open creating a wide (12 Å) entryway. Amino-acid sequence conservation suggests a common structural basis for gating in a wide range of K^+ channels, both ligand- and voltage-gated. The open conformation favours high conduction by compressing the membrane field to the selectivity filter, and also permits large organic cations and inactivation peptides to enter the pore from the intracellular solution.

Potassium and other ion channels are allosteric proteins that switch between closed and opened conformations in response to an external stimulus in a process known as gating. Depending on the channel type, the gating stimulus can be the binding of a ligand, the membrane electric field, or both. A central issue in ion channel biophysics concerns the nature of the pore conformational changes that accompany channel gating. What do the opened and closed structures of the pore look like? In general, little is known about protein conformational changes in membrane proteins, and yet for ion channels these changes are crucial to every aspect of their function: ion conduction, gating and pharmacology. Here, we address the following questions regarding these three areas of ion channel function. For the conduction mechanism of K^+ channels, when the pore opens, how wide does it become, how does opening change the electric field across the pore, and how accessible is the K^+ selectivity filter to the intracellular solution? For the gating mechanism, what are the mechanics of pore opening, are the conformational changes within the membrane large or small? Finally, for K^+ channel pharmacology, which protein surfaces become exposed when the pore opens, and how might pharmacological agents interact with the closed versus opened state of the channel?

Mutational experiments from numerous laboratories have placed important structural constraints on K^+ channel gating. In particular, mutational studies of the Shaker voltage-dependent K^+ channel^{1,2}—interpreted in the context of the KcsA K^+ channel structure³—identified what is probably the pore's gate. In the KcsA structure four α -helices (inner helices) line the pore's intracellular half, forming a right-handed bundle (inner helix bundle) near the intracellular opening (Fig. 1)^{3,4}. The inner helix bundle marks the point at which the Shaker pore becomes inaccessible to thiol-reactive compounds and metal ions applied to the cytoplasmic side of the channel when the channel is closed^{1,5}. Thus, it would seem that the extracellular half of the pore, where the selectivity filter is located, is dedicated to the function of K^+ selectivity, and that the intracellular half, where the pore is lined by inner helices, is dedicated to gating.

In the KcsA structure the pore is very wide (about 12 Å diameter) at the centre of the membrane in what is called the central cavity, but closer to the intracellular opening, at the level of the inner helix bundle, the pore diameter narrows to about 4.0 Å (the separation between van der Waals surfaces of protein atoms) (Fig. 1). At this narrow segment, hydrophobic side chains from the inner helices line the pore, creating what would seem to be an inhospitable environ-

ment for a K^+ ion. Is the gate closed in the KcsA crystal structure? Functional measurements offer the best insight into the probable conformational state. In membranes, the KcsA K^+ channel has a very low open probability even under the acidic pH conditions known to open the channel^{6,7}. When the intracellular carboxy-terminal 35 amino acids are truncated from KcsA, the open probability is even lower, remaining effectively near zero (C. Miller, personal communication). The crystal structures (Fig. 1) are of the truncated KcsA K^+ channel^{3,4}; if structure matches function, then the truncated KcsA K^+ channel has a closed gate.

Exactly how do the inner helices move to open the pore? This has been a difficult question to answer in the absence of direct measurements of closed and opened channel structures. One attempt, using electron paramagnetic resonance (EPR) with spin-labelled KcsA channels, suggested that the inner helices rotate and translate (relative to the KcsA crystal structure), causing a very subtle diameter increase at the inner helix bundle (Protein Data

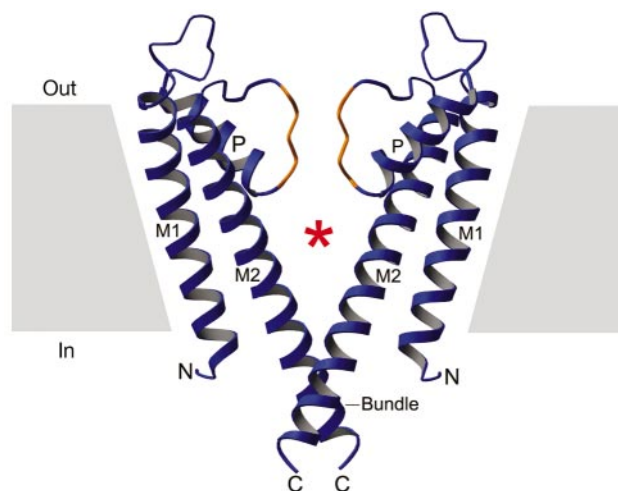


Figure 1 Structural elements of the K^+ channel pore. Two subunits of the KcsA K^+ channel are shown with the extracellular side on top. The selectivity filter is orange and the central cavity is marked by a red asterisk. Three helical segments include, from N to C terminus, the outer helix (M1), pore helix (P) and inner helix (M2). The gate is formed by the inner helix bundle (Bundle). The figure was prepared using RIBBONS²⁸.

Bank code 1JQ1)^{8,9}. However, a different approach using X-ray crystallography implied that the pore might open to a much wider diameter¹⁰. The crystallographic study showed that the K⁺ channel blocker tetrabutylammonium ion (TBA) binds in the central cavity. In the structure the inner helices remained in their apparently closed conformation, but they must have opened very wide to allow TBA (diameter 10–12 Å) to enter.

Mechanics of pore gating

The MthK channel from *Methanobacterium thermoautotrophicum* together with the KcsA channel provides valuable insight into the conformational changes that apparently underlie K⁺ channel gating. The MthK channel structure includes its gating machinery (the gating ring) and contains a bound ligand (a Ca²⁺ ion) that opens the channel in membranes¹¹ (PDB code 1LNQ). Figure 2 shows the pore structure of the MthK channel alongside the KcsA K⁺ channel (PDB code 1K4C). Comparison between these two structures revealed first that the structure surrounding the selectivity filter is quite similar in both channels, aside from differences in the length of the extracellular ‘turret’ loops. Second, there are very large structural differences involving the inner helices: in KcsA the inner helices are nearly straight and form a bundle near the

intracellular solution, whereas in MthK the inner helices are bent and splayed open (Fig. 2a–c). Third, the bend (approximately 30°) in the MthK inner helices occurs at a hinge point—a gating hinge—that is located deep within the membrane, just below the selectivity filter. The gating hinge corresponds to Gly 83 in MthK and Gly 99 in KcsA (Fig. 3). Glycine is unique in its ability to adopt a wide range of main-chain dihedral angles and confers flexibility at specific points in protein structures. The presence of a glycine at the gating hinge permits the structural differences observed between the KcsA and MthK channels. Fourth, to convert the KcsA (closed pore) structure into the MthK (opened pore) structure, it appears as though one would simply have to exert a lateral (radial-outward) force on the C-terminal (intracellular) extent of the inner helices; such a force would place a torque on the gating hinge (Fig. 2c, d). In the open conformation the contact area between the inner and outer helices is reduced, especially near the intracellular surface. This probably accounts for the straightening of the outer helices in MthK. Finally, the open pore is very wide, about 12 Å at its narrowest point (Ala 88, Fig. 3), so that the central cavity becomes essentially continuous with the intracellular solution. The structure can be viewed in the Supplementary Information as a movie.

Gating mechanism is conserved

Is it reasonable to compare two different K⁺ channels, KcsA and MthK, to infer the conformational changes that underlie gating? Structure-based sequence alignment using a wide range of K⁺ channels suggests that it is, and that the inferred conformational changes probably occur in most K⁺ channels (Fig. 3a). The K⁺ channel sequences shown in Fig. 3a cover the broad range found in nature: prokaryotic, eukaryotic, channels with two and six membrane-spanning segments per subunit, ligand-gated and voltage-gated. Even cyclic nucleotide-gated (CNG) channels, which are not K⁺ selective but are related to K⁺ channels, apparently exhibit similar conformational changes¹². The principal observation in the sequence alignment is strong amino-acid conservation at two positions in the inner helix (Fig. 3a, b). The first position is the gating hinge, conserved as glycine (red in Fig. 3a); apparently all of these channels have a gating hinge. The second position is five amino acids C-terminal to the gating hinge, conserved as alanine or glycine (green). The significance of having an amino acid with a small side chain at this second position is evident on inspection of the MthK channel structure: this amino acid points its side chain towards the pore, a large side chain would tend to plug the pore and interfere with ion conduction. Thus, the MthK structure reveals the significance of a glycine residue, to allow a flexible gating hinge, and an alanine residue, to ensure a wide pathway for ions to reach the selectivity filter. Conservation of these residues implies a conserved pore conformational change. For this reason we propose that KcsA and MthK structures represent the closed and opened states of most K⁺ channels, ligand- and voltage-gated, as well as CNG channels.

The mechanics of pore opening may be conserved in different K⁺ channels, but very different structures, such as integral membrane voltage sensors^{13,14} and intracellular ligand-binding domains¹⁵, have to mechanically adapt to the pore so that they can exert a lateral force on the inner helices. Thus, the C-terminal end of the inner helices must vary in structure from one channel to the next, depending on its associated gating domain. In the MthK channel we are unable to fully resolve the connection between the pore and its gating domain, but the inner helix points at its attachment site as if the connection is nearly straight¹¹. When the pore closes, in order for the inner helices to form a bundle—similar to the structure observed in KcsA (Fig. 2)—the connection would have to develop a bend. In this regard, much work has focused on the Pro-Val-Pro sequence near the inner helix bundle in certain voltage-dependent K⁺ channels¹. The suggestion that this sequence might make a bend seems reasonable, as a bend would direct the bottom of the inner helix towards the voltage sensor domain.

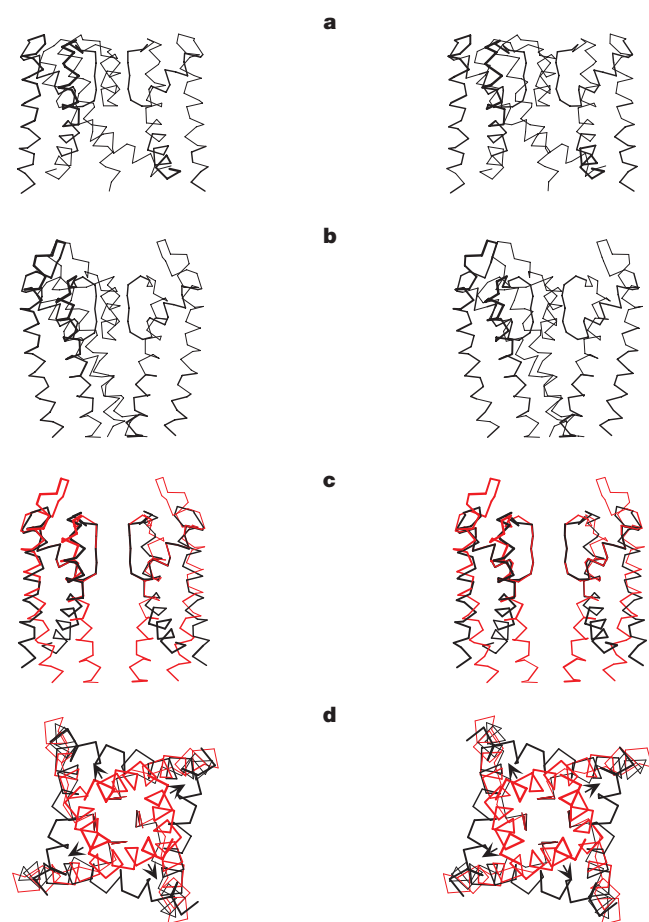


Figure 2 Opened and closed states of K⁺ channels. Stereo diagrams show the pore (Cα traces) of two K⁺ channels viewed from within the membrane: **a**, MthK (PDB code 1LNQ) (residues 19–98) and **b**, KcsA (PDB code 1K4C) (residues 24–114), with one subunit removed and oriented with the extracellular surface on top. **c**, **d**, The KcsA (red) and MthK (black) pores are superimposed and viewed from within the membrane (**c**, two subunits), and from the intracellular solution (**d**, four subunits). Only the selectivity filter and inner helices are shown in **d**. Arrows indicate the direction of inner helix displacement in going from closed (KcsA) to opened (MthK). The figures were prepared using Bobscript²⁹.

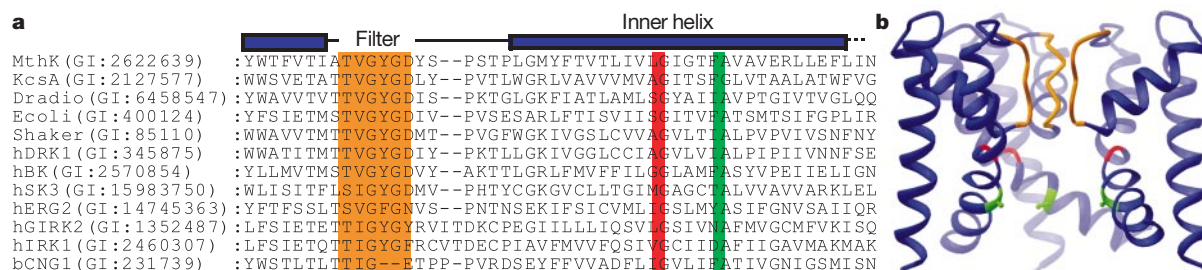


Figure 3 Structure-based sequence analysis suggests conserved gating conformations. **a**, Sequences from the inner helices of various K⁺ channels and a CNG channel. The selectivity filter is coloured orange, the gating hinge glycine red, and the amino acid at the narrowest point of the MthK intracellular pore entryway, green. MthK, *M. thermotrophicum* K⁺ channel; KcsA, *Streptomyces lividans* K⁺ channel; Dradio, *Deinococcus radiodurans* K⁺ channel; Ecoli, *Escherichia coli* K⁺ channel; Shaker,

Drosophila melanogaster K⁺ channel; hDRK1, *Homo sapiens* K⁺ channel; hBK, *H. sapiens* BK channel; hSK3, *H. sapiens* SK channel; hERG2, *H. sapiens* K⁺ channel; hGIRK2, *H. sapiens* G-protein-activated K⁺ channel 2; hIRK1, *H. sapiens* K⁺ channel; bcNG1, *Bos taurus* cyclic nucleotide-gated channel. Protein ID numbers are indicated in parentheses. **b**, Three subunits of the MthK pore with colours corresponding to **a**.

Pharmacology and inactivation gating

The wide diameter of the open gate (12 Å) is entirely consistent with the experiments of Armstrong, showing that large organic cations block the intracellular entryway of K⁺ channels^{16,17} (Fig. 4). The open gate is wide enough so that organic compounds can enter from the cytoplasm and reach the membrane centre (the cavity) before lodging just beneath the selectivity filter¹⁰. It is interesting to see that gate closure (the pinching shut of the inner helix bundle) occurs while maintaining the size of the cavity (Fig. 2). This structural feature of K⁺ channel gating is consistent with the phenomenon of functional 'trapping' of organic cations behind the gate, deep within the pore^{16–18}. The structure of the open pore builds on our understanding of the susceptibility of K⁺ channels to block by organic cations, including many commonly prescribed pharmacological agents¹⁹. On the intracellular side of the selectivity filter the pore is hydrophobic³, cation attractive (owing to oriented pore helices)²⁰ and wide (Fig. 4). These properties of K⁺ channels undoubtedly lie at the basis of certain drug-induced cardiac conduction abnormalities such as acquired long-QT syndrome¹⁹.

The structure of the open pore helps us to further understand a form of K⁺ channel gating known as inactivation. A-type K⁺ channels such as the Shaker K⁺ channel open in response to membrane depolarization, but a few milliseconds later they stop

conducting ions because an inactivation gate blocks the pore on the intracellular side^{21,22}. A 'ball and chain' mechanism was proposed about thirty years ago for a similar process in Na⁺ channels^{23,24}. Ten years ago, the 'ball and chain' inactivation gate of A-type K⁺ channels was shown to be formed by the channel's own amino terminus^{21,22}, or the N terminus of an associated β-subunit²⁵. Recently, on the basis of a mutational study, it was proposed that an N-terminal inactivation gate blocks ion conduction by entering into the pore as an extended peptide, reaching to the cavity¹⁰. The mutational data were convincing, but nevertheless, given the narrow

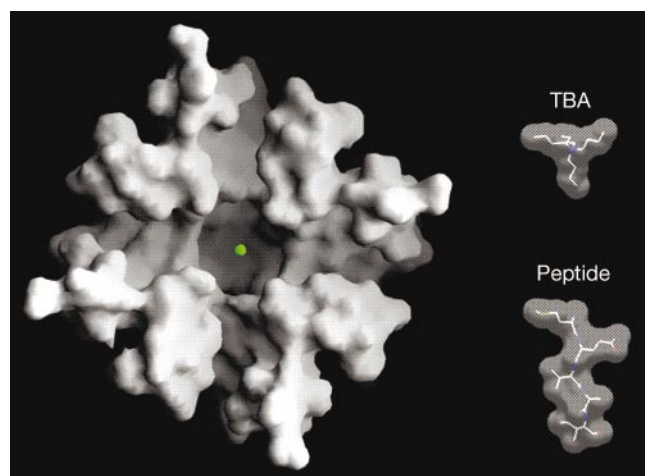


Figure 4 The open pore allows entry of large molecules from the intracellular solution. Molecular surface of the MthK pore (all side chains included) viewed from the intracellular solution. A K⁺ ion in the selectivity filter is shown as a green sphere. Surface models of tetrabutylammonium (TBA) and the N-terminal inactivation gate (sequence MQVS) from KvB1.1 (peptide) are shown¹⁰. The figure was prepared using GRASP³⁰.

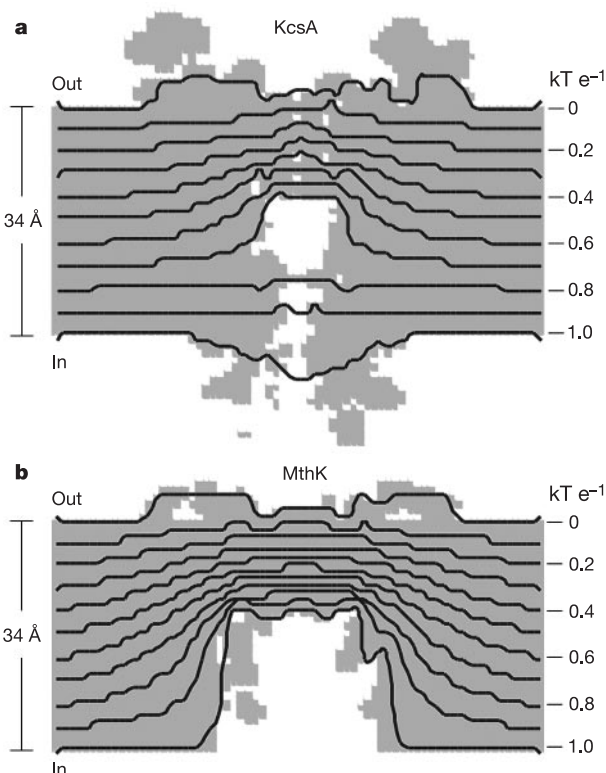


Figure 5 The membrane electric potential across the pore changes on opening. **a**, **b**, Electrostatic contour plots for KcsA (**a**) and MthK (**b**) in a membrane. Electric potential was calculated as described (see Methods). A slice perpendicular to the membrane through the channel shows the pore between the intracellular solution (bottom) and extracellular solution (top). The grey region corresponds to protein or membrane (dielectric constant 2); white regions to aqueous solution (dielectric constant 80). Potential runs from 1.0 to 0 with contours every 0.1 kT e⁻¹.

inner pore of KcsA the proposal seemed unbelievable from a structural perspective. The open channel structure of MthK, however, gives a different impression, and rationalizes the mutational results (Fig. 4). The open pore is wide enough so that an extended peptide can insert its N terminus (cationic) into the cavity.

Membrane potential across the open pore

The open pore conformation has important implications for ion conduction, in part because the electric field experienced by an ion in the pore depends on the shape of the pore, among other factors. To gain a qualitative appreciation for the effect of pore shape, we calculated the electrostatic potential for the KcsA and MthK pores, closed and opened K⁺ channels, embedded in a membrane by solving the finite difference Poisson equation. The dielectric constant for the protein and membrane were set at 2.0, and for water outside the membrane and in the pore, 80. To focus on the effect of pore shape on applied membrane potential the protein was uncharged. The results are shown in the form of electrostatic contour plots (Fig. 5). In the open conformation the cavity is at the same membrane potential as the internal solution. This Poisson calculation gives the potential under equilibrium, when ions are not moving through the pore. The qualitative result is impressive: an open K⁺ channel effectively thins the membrane for a diffusing K⁺ ion to approximately the 12 Å length of the selectivity filter. That is, the access resistance for a K⁺ ion diffusing between the cytoplasm and the selectivity filter is low. This feature of K⁺ channels undoubtedly contributes to their high conductance.

Discussion

On the basis of the opened MthK channel, the closed KcsA channel, and amino-acid sequence analysis, we propose a structural basis for gating transitions in the transmembrane pore of K⁺ channels. The four inner helices can exist in a straight (apparently relaxed) conformation in which case they form a bundle that closes the pore near its intracellular opening. Alternatively the inner helices can bend, causing the bundle to splay open to a diameter of about 12 Å. The principal structural feature of the gating conformational change is the presence of a gating hinge located deep within the membrane, conserved as a glycine residue in most K⁺ channels. The gating conformational changes within the membrane are very large, but are mainly confined to the intracellular half of the channel. Based on amino-acid sequence conservation, we propose that different K⁺ channels, ligand- and voltage-gated, as well as cyclic nucleotide-gated channels, undergo similar pore conformational changes. □

Methods

Electrostatic calculation

Electrostatic potential was calculated by solving the finite difference Poisson equation using a program written in Fortran 77 implementing a Jacobi relaxation algorithm^{26,27}. The KcsA and MthK pores were mapped onto a 70 × 70 × 90 Å cubic grid with spacings of 1.0 Å. Side chains were added to the MthK pore structure for defining the region of space bounded by the channel. Channels were placed into a 34 Å thick slab of dielectric constant 2.0 used to model the membrane. The protein region was assigned a dielectric constant of 2.0 using the standard Pauling atomic radii; if an atom intersected a grid element then a dielectric constant of 2.0 was assigned to the element. We did not use protein charges in the calculation. Solution surrounding the membrane and in the pore was assigned a dielectric constant of 80. Ionic strength outside the membrane had only a small effect on the potential within the membrane and therefore was kept at 0 for producing Fig. 5.

Received 22 March; accepted 22 April 2002.

- del Camino, D., Holmgren, M., Liu, Y. & Yellen, G. Blocker protection in the pore of a voltage-gated K⁺ channel and its structural implications. *Nature* **403**, 321–325 (2000).
- Liu, Y., Holmgren, M., Jurman, M. E. & Yellen, G. Gated access to the pore of a voltage-dependent K⁺ channel. *Neuron* **19**, 175–184 (1997).

- Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* **280**, 69–77 (1998).
- Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K⁺ channel–Fab complex at 2.0 Å resolution. *Nature* **414**, 43–48 (2001).
- del Camino, D. & Yellen, G. Tight steric closure at the intracellular activation gate of a voltage-gated K⁺ channel. *Neuron* **32**, 649–656 (2001).
- Cuello, L. G., Romero, J. G., Cortes, D. M. & Perozo, E. pH-dependent gating in the *Streptomyces lividans* K⁺ channel. *Biochemistry* **37**, 3229–3236 (1998).
- Heginbotham, L., LeMasurier, M., Kolmakova-Partensky, L. & Miller, C. Single streptomyces lividans K⁺ channels. Functional asymmetries and sidedness of proton activation. *J. Gen. Physiol.* **114**, 551–560 (1999).
- Perozo, E., Cortes, D. M. & Cuello, L. G. Structural rearrangements underlying K⁺-channel activation gating. *Science* **285**, 73–78 (1999).
- Liu, Y. S., Somporpnisut, P. & Perozo, E. Structure of the KcsA channel intracellular gate in the open state. *Nature Struct. Biol.* **8**, 883–887 (2001).
- Zhou, M., Morais-Cabral, J. H., Mann, S. & MacKinnon, R. Potassium channel receptor site for the inactivation gate and quaternary amine inhibitors. *Nature* **411**, 657–661 (2001).
- Jiang, Y. *et al.* Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417**, 515–522 (2002).
- Flynn, G. E., Johnson, J. P. Jr & Zagotta, W. N. Cyclic nucleotide-gated channels: shedding light on the opening of a channel pore. *Nature Rev. Neurosci.* **2**, 643–651 (2001).
- Sigworth, F. J. Voltage gating of ion channels. *Q. Rev. Biophys.* **27**, 1–40 (1994).
- Bezanilla, F. The voltage sensor in voltage-dependent ion channels. *Physiol. Rev.* **80**, 555–592 (2000).
- Jiang, Y., Pico, A., Cadene, M., Chait, B. T. & MacKinnon, R. Structure of the RCK domain from the *E. coli* K⁺ channel and demonstration of its presence in the human BK channel. *Neuron* **29**, 593–601 (2001).
- Armstrong, C. M. Interaction of tetraethylammonium ion derivatives with the potassium channels of giant axons. *J. Gen. Physiol.* **58**, 413–437 (1971).
- Armstrong, C. M. Ionic pores, gates, and gating currents. *Q. Rev. Biophys.* **7**, 179–210 (1974).
- Holmgren, M., Smith, P. L. & Yellen, G. Trapping of organic blockers by closing of voltage-dependent K⁺ channels: evidence for a trap door mechanism of activation gating. *J. Gen. Physiol.* **109**, 527–535 (1997).
- Mitcheson, J. S., Chen, J., Lin, M., Culberson, C. & Sanguinetti, M. C. A structural basis for drug-induced long QT syndrome. *Proc. Natl Acad. Sci. USA* **97**, 12329–12333 (2000).
- Roux, B. & MacKinnon, R. The cavity and pore helices in the KcsA K⁺ channel: electrostatic stabilization of monovalent cations. *Science* **285**, 100–102 (1999).
- Hoshi, T., Zagotta, W. N. & Aldrich, R. W. Biophysical and molecular mechanisms of Shaker potassium channel inactivation. *Science* **250**, 533–538 (1990).
- Zagotta, W. N., Hoshi, T. & Aldrich, R. W. Restoration of inactivation in mutants of Shaker potassium channels by a peptide derived from ShB. *Science* **250**, 568–571 (1990).
- Bezanilla, F. & Armstrong, C. M. Inactivation of the sodium channel. I. Sodium current experiments. *J. Gen. Physiol.* **70**, 549–566 (1977).
- Armstrong, C. M. & Bezanilla, F. Inactivation of the sodium channel. II. Gating current experiments. *J. Gen. Physiol.* **70**, 567–590 (1977).
- Rettig, J. *et al.* Inactivation properties of voltage-gated K⁺ channels altered by presence of β-subunit. *Nature* **369**, 289–294 (1994).
- Warwicker, J. & Watson, H. C. Calculation of the electric potential in the active site cleft due to alpha-helix dipoles. *J. Mol. Biol.* **157**, 671–679 (1982).
- Klapper, I., Hagstrom, R., Fine, R., Sharp, K. & Honig, B. Focusing of electric fields in the active site of Cu–Zn superoxide dismutase: effects of ionic strength and amino-acid modification. *Proteins* **1**, 47–59 (1986).
- Carson, M. Ribbons. *Methods Enzymol.* **277**, 493–505 (1997).
- Enouf, R. M. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 132–133 (1997).
- Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank the staff at the National Synchrotron Light Source, Brookhaven National Laboratory, X-25, the Cornell High Energy Synchrotron Source, F1, and the Advanced Light Source, Lawrence Berkeley Laboratory, 5.0.2 for synchrotron support; members of the MacKinnon laboratory for assistance; R. Dutzler and R. Xie for programming and graphical assistance; F. Sigworth, P. Jordan, C. Miller and D. Gadsby for discussions; and W. Chin for help in manuscript preparation. This work was supported by grants from the NIH to R.M. and from the National Center for Research Resources, NIH, to B.T.C. R.M. is an investigator in the Howard Hughes Medical Institute.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.M. (e-mail: mackinn@rockefeller.edu). Coordinates have been deposited with the Protein Data Bank under accession code 1LNQ.

Optical pulsations from the anomalous X-ray pulsar 4U0142+61

B. Kern & C. Martin

Division of Physics, Mathematics, and Astronomy, California Institute of Technology, MS 405-47, Pasadena, California 91125, USA

Anomalous X-ray pulsars¹ (AXPs) differ from ordinary radio pulsars in that their X-ray luminosity is orders of magnitude greater than their rate of rotational energy loss, and so they require an additional energy source. One possibility is that AXPs are highly magnetized neutron stars²—or ‘magnetars’—having surface magnetic fields greater than 10^{14} G. This would make them similar to the soft γ -ray repeaters (SGRs)³, but alternative models that do not require extreme magnetic fields also exist. An optical counterpart to the AXP 4U0142+61 was recently discovered⁴, consistent with emission from a magnetar, but also from a magnetized hot white dwarf⁵, or an accreting isolated⁶ neutron star⁷. Here we report the detection of optical pulsations from 4U0142+61. The pulsed fraction of optical light (27 per cent) is five to ten times greater than that of soft X-rays, from which we conclude that 4U0142+61 is a magnetar. Although this establishes a direct relationship between AXPs and the soft γ -ray repeaters, the evolutionary connection between AXPs, SGRs and radio pulsars remains controversial.

We observed the field around the faint ($R = 25$) optical counterpart that was proposed in ref. 4 for the 8.7-s AXP 4U0142+61 (Fig. 1) on 15–16 November 2001 at the Palomar Observatory Hale 5-m telescope, using a phase-binning charge-coupled device (CCD). The average seeing was 1.4 arcsec, in non-photometric conditions. The light curves of every point in the field-of-view (200×10.4 arcsec) are binned on-chip into 10 phase bins. The imaging properties of this technique allow us to ensure that the observed time variability of 4U0142+61 is not instrumental in origin. In Fig. 2 we show that the variability (at the X-ray period of 4U0142+61) everywhere but the optical counterpart is consistent with the noise. The light curve of 4U0142+61 is shown Fig. 3.

The remarkable aspect of the optical light curve is that its modulation amplitude is very large compared to that of the X-ray light curves. We define pulsed fraction h as

$$h = \frac{F_{\max} - F_{\min}}{F_{\max} + F_{\min}} \quad (1)$$

where $F_{\max}$ and $F_{\min}$ are the maximum and minimum flux in any phase bin. We use a maximum-likelihood estimator for h to reduce the systematic bias caused by random errors. We find $h_{\text{opt}} = 27^{+8}_{-6}\%$ (1σ errors), with a probability of spurious detection of 6×10^{-5} (one-tailed). Owing to the finite temporal resolution of the optical data, this estimate is a lower limit to the true pulsed fraction.

The X-ray spectrum of 4U0142+61 is generally interpreted as an absorbed sum of power-law and blackbody components⁸, with a power-law index of around 3.7 and temperature 0.4 keV. With this ultrasoft power-law index, the observed luminosity, corrected for absorption, is dominated by soft X-rays. Between 0.5 and 10.0 keV, the 0.5–1.0, 1.0–2.0 and 2.0–10.0 keV X-rays provide 65%, 25% and 10% of the unabsorbed flux, respectively. To compare the optical and soft X-ray pulse profiles, we use archival Chandra ACIS-S (Advanced CCD Imaging Spectrometer Spectroscopic array) data (Fig. 3), which provide the most sensitivity to 0.5–1.0 keV X-rays of any observations to date. Following equation (1) to define $h_{0.5}$ and $h_{1.0}$ as the pulsed fractions for 0.5–1.0 and 1.0–2.0 keV X-rays, respectively, we find $h_{0.5} = 2.9^{+1.9}_{-2.9}\%$ and $h_{1.0} = 5.0 \pm 0.5\%$. The

quoted errors are 1σ errors, which show that $h_{0.5}$ is consistent with zero. For these estimates, the X-ray data were binned into 10 phase bins to match the time resolution of the optical data. Binning the 1.0–2.0 keV X-rays into 20 phase bins instead of 10 increases the measured pulsed fraction by 2%.

We first consider the optical pulsations in the framework of the magnetar model. There is compelling evidence that the soft γ -ray repeaters are magnetars³, but the SGRs are heavily absorbed at visible wavelengths, and there are no detailed models predicting the optical emission from magnetospheres of magnetars. We turn to the closest relatives of magnetars, the young (10^3 – 10^4 yr) isolated radio pulsars with measured optical pulsations, that is, Crab^{9,10}, Vela^{11,12} and PSR B0540-69^{13,14}. The optical light curves for these pulsars are similar in morphology to their X-ray light curves, and they have $L_X/L_{\text{opt}} \approx 10^3$ – 10^4 , similar to 4U0142+61. At present, the data are consistent with the optical emission that might be expected from magnetars.

We next consider the possibility that 4U0142+61 is a neutron star fed by accretion, without a stellar companion⁶. We assume that beamed X-rays are emitted at the neutron star surface, and that rotation of the neutron star leads to the observed X-ray pulsations. Optical pulsations at the X-ray pulsation frequency can arise only from thermal reprocessing of X-rays illuminating a disk. In binary systems, where the X-rays illuminate a companion star, the pulsed fraction of the reprocessed light cannot exceed the pulsed fraction of the incident X-rays¹⁵. The reprocessing mechanism is similar in accretion disks¹⁶, giving

$$h_{\text{opt}} \leq h_X \quad (2)$$

where h_X is the pulsed fraction of the X-rays illuminating the disk. Assuming that the softest X-rays (0.5–1.0 keV) dominate the energetics of reprocessing and that the X-rays illuminating the disk are the same as the observed X-rays (that is, $h_X = h_{0.5}$), that is violated, because $h_{\text{opt}} \approx 10h_{0.5}$.

To reconcile equation (2) with the observations, two conditions must be satisfied; $h_X > h_{0.5}$, and the disk axis and the neutron star rotation axis must be misaligned, so that the X-ray illumination of the disk is not azimuthally symmetric (or else there would be no optical pulsations). The first condition, $h_X > h_{0.5}$, could be satisfied

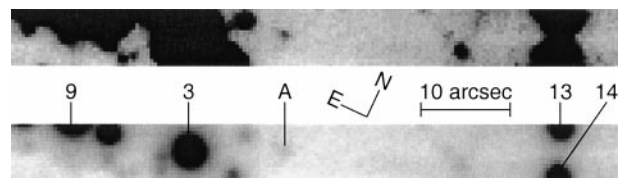


Figure 1 Time-averaged optical image of field around 4U0142+61. This image is 170×16 pixels (68×6.4 arcsec), displayed with linear (top panel) and logarithmic (bottom panel) scaling. The phase-binning CCD has 512×512 pixels, masked by a 500×26 pixel (200×10.4 arcsec) slit. Intensity is binned on-chip into ten phase intervals using a periodic frame transfer of the CCD. Every $1/10$ of the period of 4U0142+61 (known *a priori*), the accumulated image is transferred away from the slit onto a region of the CCD used for storage, but no pixels are read out of the CCD. At the end of ten transfers (that is, 1 period of 4U0142+61), which alternately shift the accumulated charge up and down on the CCD, the image accumulated during the first phase interval is shifted back into the CCD region illuminated by the slit, and the image is re-exposed. At any time, there are ten accumulated images residing on the CCD, one of which is being exposed. This exposure pattern is repeated for 95 s (11 cycles of 4U0142+61), at which point the shutter is closed and the entire array is read out. Each readout delivers ten phase-binned images, each of which has been exposed for 9.5 s, long enough for the read noise of the CCD to be small compared to the Poisson noise of the sky background. The total exposure time in two nights was 60,000 s. No filter was used, giving a spectral response of 400–1,000 nm. The images shown here are sums of all ten phase-binned images. The numbering of stars follows the convention of refs 28 and 29 with 4U0142+61 labelled star A, as in ref. 4.

if the X-rays were more strongly beamed toward the disk than toward the observer. This is unlikely to be significant, because general relativistic effects spread out beamed X-rays from the neutron star surface¹⁷. For a $1.4M_{\odot}$ neutron star with a 10-km radius, where $M_{\odot}$ is the mass of the Sun, the flux ratio between the centre of the X-ray beam and a line-of-sight 90° away will be less than 4 (ref. 18). Another reason h_X may be greater than $h_{0.5}$ is that the pulsed fraction of X-rays with energies less than 0.5 keV (unobservable because of absorption) may be higher than $h_{0.5}$. Although there is no observational evidence for or against this possibility, we note that the trend of pulsed fraction with energy is to lower pulsed fractions with lower energies⁸.

All of the preceding neutron star discussions relate the X-rays to the reprocessed optical flux. The V-, R- and I-band optical fluxes measured in ref. 4 limit the possible size of an accretion disk to $R_{\text{out}} < 0.05R_{\odot}$, where $R_{\odot}$ is the radius of the Sun. For such a small disk, the disk heating is expected to be dominated by viscous dissipation rather than by X-ray irradiation¹⁹. The optical flux is then dominated by a steady component owing to the viscous dissipation. For these reasons, we conclude that our measurement of optical pulsations is inconsistent with the neutron star accretion model.

The third model we consider is that 4U0142+61 is a $B = 5 \times 10^8$ G, $T = 4 \times 10^5$ K white dwarf, possibly the result of a double-degenerate merger⁴. The white dwarf with the most similar physical properties is RE J0317–853, with $B = 5 \times 10^8$ G, $T = 5 \times 10^4$ K and $P = 725$ s (refs 20, 21). RE J0317–853 shows optical pulsations with $h_{\text{opt}} \approx 10\%$ (ref. 20) and 100-Å extreme ultraviolet pulsations with $h_{\text{EUV}} \approx 20\%$ (ref. 21). Two other highly magnetized, rotating white dwarfs have shown photometric optical variability: PG1031+234 with $h \approx 20\%$, 15% and 10% for U, B and V (ref. 22), and Feige 7 with $h \approx 5\%$ in V (ref. 23). Surface variations in magnetic field strength in these objects change the continuum opacity (magnetic dichroism), and may cause variations in surface temperature. In addition, the surface abundances may be nonuniform. Models have successfully explained the photometric variability of these white dwarfs by adjusting their geometric and chemical parameters.

We know of no model atmospheres that accommodate both the high temperatures and large magnetic fields proposed for 4U0142+61. Model atmospheres with temperatures of 10^5 – 10^6 K (but no magnetic fields) have X-ray spectra that are very sensitive to both temperature and chemical abundance. Assuming that the observed X-rays are related to the extreme ultraviolet/soft X-rays emitted near the peak of the blackbody spectrum, the X-rays scale as T^4 , while the optical luminosity scales as T^1 , giving $h_X > h_{\text{opt}}$. Even in a pure hydrogen atmosphere, there is a temperature-dependent photoionization opacity that will cause more variability in the X-ray fluxes²⁴. Ionization edges of highly ionized He, C, N, O and Ne are very sensitive to changes in temperature, and line blanketing by these elements results in extreme ultraviolet/soft X-ray spectra that vary by orders of magnitude for temperature differences of a factor of 2 (ref. 25). We conclude that the mechanisms that have explained

optical photometric variability in other white dwarfs are likely to produce $h_{\text{opt}} < h_X$ (as in RE J0317–853, with $h_{\text{opt}} = h_{\text{EUV}}/2$).

The X-ray spectrum of 4U0142+61 is difficult to explain in the context of a white dwarf. If the blackbody component is to be believed, then the emission region is confined to a radius of 12 km (ref. 8; assuming a distance of 5 kpc). Assuming $M \approx 1.3M_{\odot}$, and $R \approx 0.01R_{\odot}$, the surface gravity is $\log(g) \approx 9$, which means that, assuming the blackbody radiation is emitted at the surface, the flux in the emitting region is approximately 16 times the Eddington flux.

The evidence that 4U0142+61 is a magnetar is strengthened by the disagreement between the accretion and white dwarf models and the optical observations. Optical observations of magnetars open up new avenues for exploring exotic effects, such as quantum electrodynamics in strong magnetic fields resulting in magnetic photon splitting²⁶ and polarization of the vacuum index of refraction²⁷. High-quality pulse-phased photopolarimetry is possible with 8–10 m telescopes, opening up a new field of physical exploration. Future multi-colour optical and infrared phase, morphology, and polarization observations, when combined with X-ray data, will provide crucial diagnostics of magnetar physics in much the same way that multi-wavelength pulse-phased observations have proved invaluable for expanding our understanding of radio pulsars. In

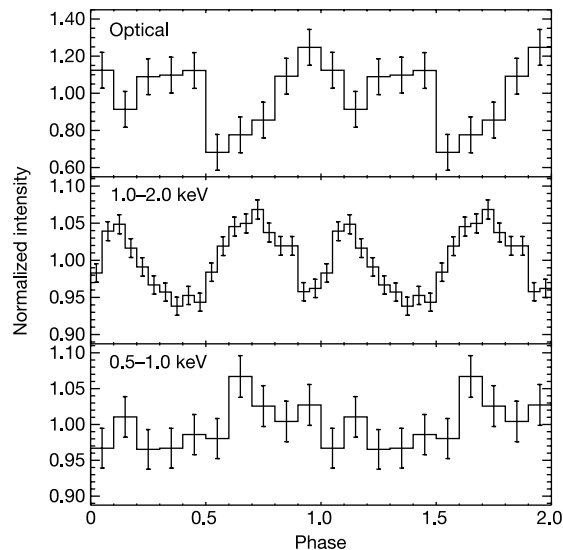


Figure 3 Optical and X-ray pulse profiles of 4U0142+61. Each panel is normalized to the mean flux at that energy, with 1σ error bars. The optical flux is measured in a 4×4 -pixel (1.6×1.6 -arcsec) aperture centred on 4U0142+61. The timing information used for these exposures was taken from the “second span” entry in Table 2 of ref. 30, with an arbitrary choice of absolute zero-phase. These parameters are: $\nu = 0.1150969336$ Hz, $\dot{\nu} = -2.687 \times 10^{-14}$ Hz s^{−1}, on modified Julian date (MJD) 51704.0 TDB, with zero-phase barycentric time-of-arrival MJD 51704.00006816 TDB. The timing of the optical observations is controlled by a global positioning system (GPS) receiver, that delivers start and stop pulses accurate to 1 μ s. The Crab pulsar was observed each night to verify the accuracy of the timing. The middle and lower panels are pulse profiles of 1.0–2.0 keV and 0.5–1.0 keV X-rays, respectively, with the intensity axis zoomed by a factor of 4 relative to the optical pulse profile panel. The X-ray pulse profiles are 5,000 s of Chandra ACIS-S continuously clocked data taken on 21 May 2000, folded onto the same timing solution as the optical data. The background level is less than 10^{-3} the total flux level in both plots. We do not correct the continuously clocked data for dither, which results in periodic timing errors of 50 ms ($0.006/\nu$). The absolute accuracy of Chandra continuously clocked timing has been estimated at less than 20 ms. The timing solution of ref. 30 is from RXTE data up to June 2001, five months before the Palomar observations. Because the X-ray and optical data are not contemporaneous, and the optical observations were taken outside the span of the timing solution used, the relative phases are unconstrained.

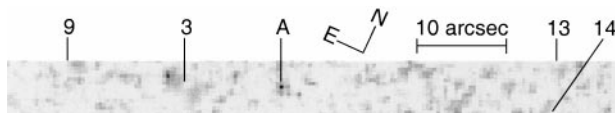


Figure 2 Map of pulsation amplitudes. At each point of the field in Fig. 1, the square of the intensity fluctuations divided by the variance is summed over all ten phase bins. Each point in this map shows the probability of chance occurrence of the residuals (plotted logarithmically). In a field this large (68×6.4 arcsec) we expect residuals as large as the spurious peak (2 arcsec east of star 3) 24% of the time. The residuals in 4U0142+61 have a chance probability of 6×10^{-5} .

addition to the physical properties of magnetars themselves, anomalous X-ray pulsars may share an evolutionary connection with soft γ -ray repeaters and radio pulsars, and could represent a significant fraction of all stellar remnants. Establishing 4U0142+61 as a magnetar is the first step in this investigation. □

Received 13 December 2001; accepted 12 April 2002.

1. Mereghetti, S. in *The Neutron Star–Black Hole Connection* (eds Kouveliotou, C., Ventura, J. & van den Heuvel, E. P. J.) 351–368 (Kluwer Academic, Dordrecht, 2001); also preprint astro-ph/9911252 at xxx.lanl.gov (1999).
2. Duncan, R. C. & Thompson, C. Formation of very strongly magnetized neutron stars: Implications for gamma-ray bursts. *Astrophys. J.* **392**, L9–L13 (1992).
3. Kouveliotou, C. et al. An X-ray pulsar with a superstrong magnetic field in the soft γ -ray repeater SGR 1806–20. *Nature* **393**, 235–237 (1998).
4. Hulleman, F., van Kerkwijk, M. H. & Kulkarni, S. R. An optical counterpart to the anomalous X-ray pulsar 4U0142+61. *Nature* **408**, 689–692 (2000).
5. Paczyński, B. X-ray pulsar 1E 2259+586: A merged white dwarf with a 7 second rotation period? *Astrophys. J.* **365**, L9–L12 (1990).
6. Wilson, C. A., Dieters, S., Finger, M. H., Scott, D. M. & van Paradijs, J. *Rossi X-ray Timing Explorer* observations of the anomalous pulsar 4U 0142+61. *Astrophys. J.* **513**, 464–470 (1999).
7. Marsden, D., Lingenfelter, R. E., Rothschild, R. E. & Higdon, J. C. Nature versus nurture: the origin of soft gamma-ray repeaters and anomalous X-ray pulsars. *Astrophys. J.* **550**, 397–409 (2001).
8. White, N. E., Angelini, L., Ebisawa, K., Tanaka, Y. & Ghosh, P. The spectrum of the 8.7 s X-ray pulsar 4U 0142+61. *Astrophys. J.* **463**, L83–L86 (1996).
9. Romani, R. W., Miller, A. J., Cabrera, B., Nam, S. W. & Martinis, J. M. Phase-resolved Crab studies with a cryogenic transition-edge sensor spectrophotometer. *Astrophys. J.* **563**, 221–228 (2001).
10. Tennant, A. F. et al. Discovery of X-ray emission from the Crab pulsar at pulse minimum. *Astrophys. J.* **554**, L173–L176 (2001).
11. Helfand, D. J., Gotthelf, E. V. & Halpern, J. P. Vela pulsar and its synchrotron nebula. *Astrophys. J.* **556**, 380–391 (2001).
12. Manchester, R. N., Wallace, P. T., Peterson, B. A. & Elliott, K. H. The optical pulse profile of the Vela pulsar. *Mon. Not. R. Astron. Soc.* **190**, 9P–11P (1980).
13. Gotthelf, E. V. & Wang, Q. D. A spatially resolved plerionic X-ray nebula around PSR B0540–69. *Astrophys. J.* **532**, L117–L120 (2000).
14. Boyd, P. T. et al. High-Speed Photometer observations of the LMC pulsar B0540–69. *Astrophys. J.* **448**, 365–368 (1995).
15. Chester, T. J. Continuum optical pulsation from the companions of binary X-ray pulsars. *Astrophys. J.* **227**, 569–578 (1979).
16. Pedersen, H. et al. Simultaneous optical and X-ray bursts from 4U/MXB 1636–53. *Astrophys. J.* **263**, 325–339 (1982).
17. Pechenick, K. R., Ftacals, C. & Cohen, J. M. Hot spots on neutron stars: The near-field gravitational lens. *Astrophys. J.* **274**, 846–857 (1983).
18. Perna, R. & Hernquist, L. Disks irradiated by beamed radiation from compact objects. *Astrophys. J.* **544**, L57–L60 (2000).
19. Hulleman, F., van Kerkwijk, M. H., Verbunt, F. W. M. & Kulkarni, S. R. A deep search for the optical counterpart to the anomalous X-ray pulsar 1E 2259+58.6. *Astron. Astrophys.* **358**, 605–611 (2000).
20. Barstow, M. A. et al. RE J0317–853: the hottest known highly magnetic DA white dwarf. *Mon. Not. R. Astron. Soc.* **277**, 971–985 (1995).
21. Ferrario, L., Vennes, S., Wickramasinghe, D. T., Bailey, J. A. & Christian, D. J. EUVE J0317–855: a rapidly rotating, high-field magnetic white dwarf. *Mon. Not. R. Astron. Soc.* **292**, 205–217 (1997).
22. Pirola, V. & Reiz, A. The highly magnetic ($B \approx 500$ MG) white dwarf PG 1031+234: Discovery of wavelength dependent polarization and intensity variations. *Astron. Astrophys.* **259**, 143–148 (1992).
23. Achilleos, N., Wickramasinghe, D. T., Liebert, J., Saffer, R. A. & Grauer, A. D. Exploring the peculiar magnetic field of Feige 7. *Astrophys. J.* **396**, 273–288 (1992).
24. Heise, J., van Teeseling, A. & Kahabka, P. The spectra and luminosity of super-soft X-ray sources. *Astron. Astrophys.* **288**, L45–L48 (1994).
25. Rauch, T. in *Supersoft X-ray Sources* (ed. Greiner, J.) 139–142 (Springer, Berlin, 1996).
26. Baring, M. G. & Harding, A. K. Photon splitting and pair creation in highly magnetized pulsars. *Astrophys. J.* **547**, 929–948 (2001).
27. Shaviv, N. J., Heyl, J. S. & Lithwick, Y. Magnetic lensing near ultramagnetized neutron stars. *Mon. Not. R. Astron. Soc.* **306**, 333–347 (1999).
28. Reid, C. A. et al. Unidentified galactic X-ray sources positioned with the HEAO-1 scanning modulation collimator. *Astron. J.* **85**, 1062–1070 (1980).
29. Coe, M. J. & Pightling, S. L. Deep infrared and optical studies of the fields of three proposed TZO remnants. *Mon. Not. R. Astron. Soc.* **299**, 223–228 (1998).
30. Gavril, F. P. & Kaspi, V. M. Long-term *Rossi X-ray Timing Explorer* monitoring of anomalous X-ray pulsars. *Astrophys. J.* **567**, 1067–1076 (2002).

Acknowledgements

We thank S. Kaye, the Palomar Observatory staff, D. Kaplan, S. Kulkarni and R. Rutledge for assistance. This work was supported by grants from the National Science Foundation and the California Institute of Technology.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.K. (e-mail: bdk@srl.caltech.edu).

Atom–molecule coherence in a Bose–Einstein condensate

Elizabeth A. Donley, Neil R. Claussen, Sarah T. Thompson & Carl E. Wieman

JILA, University of Colorado and National Institute of Standards and Technology, Boulder, Colorado 80309-0440, USA

Recent advances in the precise control of ultracold atomic systems have led to the realisation of Bose–Einstein condensates (BECs) and degenerate Fermi gases. An important challenge is to extend this level of control to more complicated molecular systems. One route for producing ultracold molecules is to form them from the atoms in a BEC. For example, a two-photon stimulated Raman transition in a ^{87}Rb BEC has been used to produce $^{87}\text{Rb}_2$ molecules in a single rotational-vibrational state¹, and ultracold molecules have also been formed² through photo-association of a sodium BEC. Although the coherence properties of such systems have not hitherto been probed, the prospect of creating a superposition of atomic and molecular condensates has initiated much theoretical work^{3–7}. Here we make use of a time-varying magnetic field near a Feshbach resonance^{8–12} to produce coherent coupling between atoms and molecules in a ^{85}Rb BEC. A mixture of atomic and molecular states is created and probed by sudden changes in the magnetic field, which lead to oscillations in the number of atoms that remain in the condensate. The oscillation frequency, measured over a large range of magnetic fields, is in excellent agreement with the theoretical molecular binding energy, indicating that we have created a quantum superposition of atoms and diatomic molecules—two chemically different species.

A Feshbach resonance is a scattering resonance for which the total energy of two colliding atoms is equal to the energy of a bound molecular state, and atom–molecule transitions can occur during a collision^{8–12}. A schematic representation of the potentials involved is shown in the inset of Fig. 1a. For our ^{85}Rb resonance, BEC atoms in the $F = 2$, $m_F = -2$ state collide on the open-channel threshold. F and m_F are the total spin and spin-projection quantum numbers. The bound state in the closed channel differs in energy by an amount ϵ from the open-channel threshold. The bound molecular wavefunction can be described as a sum of amplitudes of different hyperfine components (F, m_F) having $M_F = m_{F1} + m_{F2} = -4$ (ref. 13). Because of their different spin configurations, the atoms and molecules generally have different magnetic moments and the difference depends on magnetic field. Thus ϵ depends on magnetic field and the degree of atom–molecule coupling is magnetically tunable. The energy difference between the free atoms and the bound molecules is plotted in Fig. 1a. This behaviour of the bound-state energy also causes a resonance in the scattering length, a , which is shown in Fig. 1b. The scattering length characterizes the mean-field interaction energy of a BEC. When the magnetic field is tuned to values near the Feshbach resonance, theory predicts coherent oscillations between the atomic and molecular states, but there is significant disagreement on the conversion fraction and the coherence properties^{14–19}.

In experiments with a sodium BEC²⁰, inelastic losses were greatly enhanced when the magnetic field was ramped across the Feshbach resonance. We observed similar results for ^{85}Rb , but with lower rates^{21,22}. It is likely that the formation of molecules played a role in the atom loss, but there was no experimental evidence for the presence of molecules and the results followed a loss-rate dependence on time. More recently, we measured the time dependence of the loss in an ^{85}Rb BEC by applying controlled magnetic-field pulses

toward but not across the Feshbach resonance²³. We observed the surprising result that under some conditions, shorter, more rapid pulses actually led to more loss than longer, slower pulses that spent more time near the resonance. The time dependence of the loss was suggestive of a nonadiabatic mixing of states, with the only states within a reasonable energy range being the normal atomic BEC state and the nearby bound molecular state.

Here we show that much of the loss is probably due to the coherent mixing of atomic and molecular states. To create a superposition and probe its coherence, we applied two short magnetic-field pulses toward the Feshbach resonance, separated by a 'free evolution' time during which the magnetic field was held at a constant value some distance from the resonance. We measured the number of atoms in the condensate as a function of time between the two pulses for various values of the steady-state magnetic field between the pulses. We observed dramatic oscillations in the number of atoms remaining in the atomic BEC at frequencies corresponding to the energy splitting between the molecular and the atomic states.

The apparatus has been described in detail elsewhere^{21,23}. We first created ⁸⁵Rb condensates²¹ typically containing 16,500 atoms, with fewer than 1,000 uncondensed thermal atoms. The initial number N_{init} fluctuated from shot to shot by about 500 atoms (~3% number noise). After producing the condensate at a field of ~162 G, we ramped the magnetic field adiabatically to ~166 G, corresponding to an initial scattering length $a_{\text{init}} \approx 10a_0$, where $a_0 = 0.053$ nm. The spatial distribution of the atoms was gaussian with a peak atom density of $n_0 = 5.4 \times 10^{13} \text{ cm}^{-3}$, and the trap

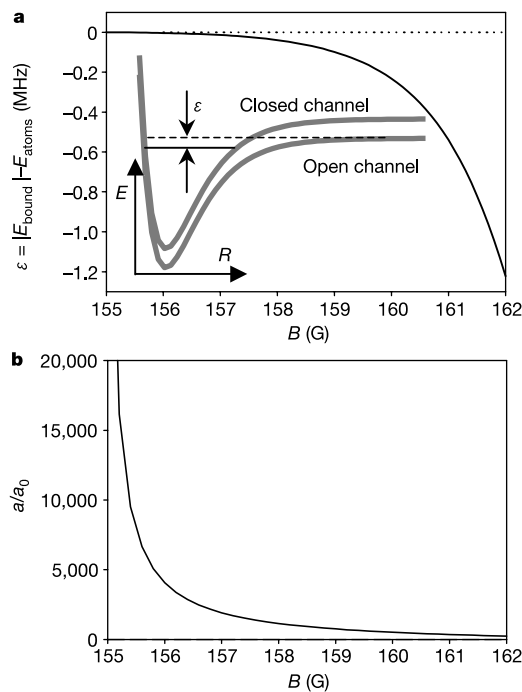


Figure 1 Feshbach resonance bound-state energy and scattering length. **a**, Energy splitting versus magnetic field. The resonance, B , is centred at ~155 G. The solid curve is a theoretical estimate of the energy found with a coupled-channels calculation, and the dotted line indicates $\varepsilon = 0$. The coupled-channels calculations were provided by S. J. J. M. F. Kokkelmans and C. H. Greene. The calculations used the best estimates of the Feshbach resonance parameters found by combining the results of three high-precision experiments. The determination of the parameters and the construction of the potentials is described in detail in ref. 29. The inset schematically shows the collision channels involved in the resonance. ε depends on magnetic field because the atoms and molecules have different magnetic moments and thus the potentials have different Zeeman shifts. **b**, Scattering length versus magnetic field for fields above the Feshbach resonance.

frequencies were $(17.5 \times 17.2 \times 6.8)$ Hz. After preparing the condensate we applied a selected fast magnetic-field pulse sequence by sending an appropriate time-dependent current through an auxiliary magnetic-field coil²³. A typical pulse sequence is shown in Fig. 2. It is composed of two nearly identical short trapezoidal pulses separated by a region of constant (but adjustable) magnetic field. Upon completion of the fast-pulse sequence in Fig. 2, we ramped the magnetic field from ~166 G to ~157 G in 5 ms and held at that field for an additional 7 ms to allow the repulsive mean-field energy to expand the condensate. Then we turned off the magnetic trap and used destructive absorption imaging 12.8 ms later to observe the atomic condensate and measure the number of remaining atoms²⁴. This detection scheme was sensitive neither to atoms with kinetic energies larger than ~2 μ K nor to atoms in off-resonant molecular states. We determined the value of the magnetic field between the pulses, B_{evolve} , by measuring the resonance frequency for transitions from the $F = 2$, $m_F = -2$ to the $F = 2$, $m_F = -1$ spin state by applying a 10- μ s radio-frequency pulse to a trapped cloud of atoms²³.

As we observed for single pulses toward the Feshbach resonance²³, there were two distinct components of atoms present in the absorption images and a third 'missing' component that we could not detect. One of the observed components was a cold remnant BEC which was not noticeably heated or excited by the fast-pulse sequence, while the other component was a relatively hot (~150 nK) 'burst' of atoms that remained magnetically trapped during the BEC expansion time. Using a variational approach²⁵ to model the mean-field expansion that we applied to the BEC remnant to measure its number, N_{remnant} , we found that we should impart at most 3 nK worth of energy to the remnant before imaging. This estimate agrees well with the expansion velocity that we observed after the trap turn-off. Thus the remnant BEC was approximately 50 times colder than the burst.

The missing component contained atoms that were in the initial sample but were not detected after the trap turn-off. To find the number of atoms in the remnant BEC and the number of burst atoms, we allowed the magnetic trap to 'focus' the burst cloud before imaging²⁶. A typical image is shown in Fig. 3. We fitted the focused burst (which had a much larger spatial extent than the remnant) with a two-dimensional gaussian surface, excluding the central region of the image that contained the remnant. This fit yielded the number of burst atoms, N_{burst} . Subtracting this fit from the image and performing a pixel-by-pixel sum of the central region of the image gave N_{remnant} .

N_{remnant} versus t_{evolve} is plotted in Fig. 4 for two different values of B_{evolve} . The number clearly oscillates. Changing the value of B_{evolve} strongly affected the oscillation frequency (note the change in scale

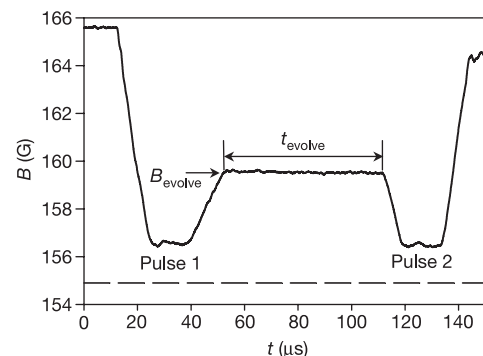


Figure 2 Magnetic field pulse shape. Fields shown for pulses 1 and 2 correspond to scattering lengths of ~2,500 a_0 , and the free precession field B_{evolve} corresponds to a scattering length of ~570 a_0 . The dashed line indicates the position of the Feshbach resonance. In the text, we refer to the free precession time as t_{evolve} . The average rise/fall time for all of the pulses that we used was 14 μ s.

from Fig. 4a to Fig. 4b). After only pulse 1 and the subsequent constant field but with no pulse 2, N_{remnant} showed no variation except for a slow decay consistent with the loss rate expected for a single pulse to that field²³.

We have taken data similar to those in Fig. 4 for a variety of different B_{evolve} values. As in Fig. 4, we fitted each curve to the function $y = y_0 + A \exp(-t/\tau_{\text{decay}}) \sin(2\pi\nu t + \phi)$ to find the oscillation frequency ν and decay time constant τ_{decay} . The measured frequencies are plotted versus B_{evolve} in Fig. 5 along with theoretical predictions for the bound-state energy relative to the atomic state.

In the regime where the scattering length is much larger than the radius of the interatomic potential well, the bound state energy for an arbitrary attractive potential can be approximated by²⁷ $\varepsilon = -\hbar^2/ma^2$. $\hbar$ is Planck's constant divided by 2π , m is the atomic mass, and a is the scattering length. The same equation relates the bound state energy to the effective scattering length, which is calculated from the Feshbach resonance parameters through the relation $a = a_{\text{bg}} \times (1 - \Delta/(B - B_0))$, where a_{bg} is the background scattering length, Δ is the width of the Feshbach resonance, and B_0 is the resonant magnetic field. Rethermalization measurements in ⁸⁵Rb thermal clouds²⁸ yield values of $B_0 = 154.9(4)$ G and $\Delta = 11.0(4)$ G. The best estimate of a_{bg} is an estimate found by mass scaling (S.J.J.M.F. Kokkelmans, personal communication) spectroscopic measurements²⁹ of the four highest-lying bound states of ⁸⁷Rb, which gives $a_{\text{bg}} = -450(3)a_0$.

The quantity $|e|/\hbar$ is plotted with no adjustable parameters in Fig. 5. The measured oscillation frequencies are in excellent agreement with this simple model over the range of magnetic fields where the model is expected to be valid. The theoretical results found with a much more sophisticated coupled-channels scattering calculation in Fig. 5 are in excellent agreement with the data over the entire range.

The fact that the oscillations occurred at exactly the frequency corresponding to the bound-state energy clearly indicates that we are creating a coherent superposition of atoms and molecules with the sudden magnetic-field pulses^{11,12,14–19}. Although we do not have a detailed understanding of how the field pulses couple atoms and molecules, by choosing the shapes of the perturbing pulses such that a single pulse results in roughly 50% atom loss, we observe high-contrast oscillations in the number of atoms in the atomic BEC.

From the amplitude of the oscillations, we can put a lower bound on the number of molecules being created. Take, for example, the data in Fig. 4a. The amplitude of the atom oscillations was 1,800(300) atoms. Assuming the fringes are coming from interference with molecules, there must be at least $1,800/2 = 900(200)$ molecules on average. Assuming that the missing atoms are molecules that we fail to convert back into atoms gives an upper bound of 3,200(100) molecules for the conditions of Fig. 4a.

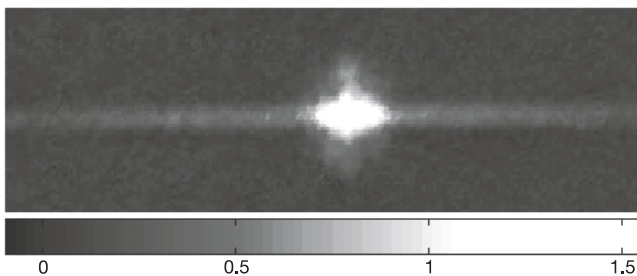


Figure 3 An absorption image taken after the fast magnetic-field pulse sequence and the mean-field expansion. The shaded bar indicates the optical density. The horizontal and vertical directions coincide with the axial and radial axes of the trap, respectively. The dimensions of the image are $366 \times 52 \mu\text{m}$. The BEC remnant is the roughly spherical cloud at the centre, while the burst atoms are focused into a thin line along the axial direction. We note the dramatic difference between the two spatial distributions, owing to the large difference in their mean energies: $\langle E_{\text{burst}} \rangle \approx 50 \times \langle E_{\text{remnant}} \rangle$.

The damping time for the oscillations, τ_{decay} , was more difficult to measure with high precision than the oscillation frequency. To within our measurement precision, τ_{decay} did not depend on B_{evolve} , but our uncertainties in τ_{decay} were as large as 100% for some fields. We had the highest precision measurements for frequencies around 200 kHz where the oscillation period was long compared to our experimental timing jitter but short compared to τ_{decay} . At frequencies near 200 kHz, we measured $\tau_{\text{decay}} = 38(8) \mu\text{s}$ for $n_0 = 1.3 \times 10^{14} \text{ cm}^{-3}$ and $\tau_{\text{decay}} = 91(33) \mu\text{s}$ when we decreased n_0 to $1.1 \times 10^{13} \text{ cm}^{-3}$.

N_{burst} also had interesting dependencies. For the conditions under which most of the data were collected, the burst contained $\sim 5,000$ atoms on average, which is $\sim 30\%$ of N_{init} . N_{burst} depended on density and varied from one-half of the atoms lost from the condensate for our typical peak density of $n_0 = 5.4 \times 10^{13} \text{ cm}^{-3}$ to nearly all of the atoms lost from the condensate for $n_0 = 1.1 \times 10^{13} \text{ cm}^{-3}$. N_{burst} , N_{remnant} , and total number of atoms detected are plotted in Fig. 6 for $B_{\text{evolve}} = 159.84(2)$ G and $n_0 = 1.1 \times 10^{13} \text{ cm}^{-3}$ (5 times lower density than was used for the data shown in Fig. 4). All three components oscillated at the same frequency. The burst oscillation lagged behind the remnant oscillation by $155(4)^\circ$. The relative phase shift was nearly 180° , so the oscillation amplitude for the total number was smaller than either the burst or the remnant oscillation amplitudes. The relative phase depended sensitively on the fall time of pulse 2. For example, when we increased the fall time from $11 \mu\text{s}$ to $159 \mu\text{s}$, the burst oscillation then lagged behind the remnant oscillation by $68(7)^\circ$ and the peak-to-peak amplitude of the total number oscillation was $5,600(400)$.

For the conditions of Fig. 6, $N_{\text{init}} = 17,100$ exceeded the time-averaged total number of atoms counted after the pulse sequence by $8(3)\%$ on average. For the higher-density measurements in Fig. 4,

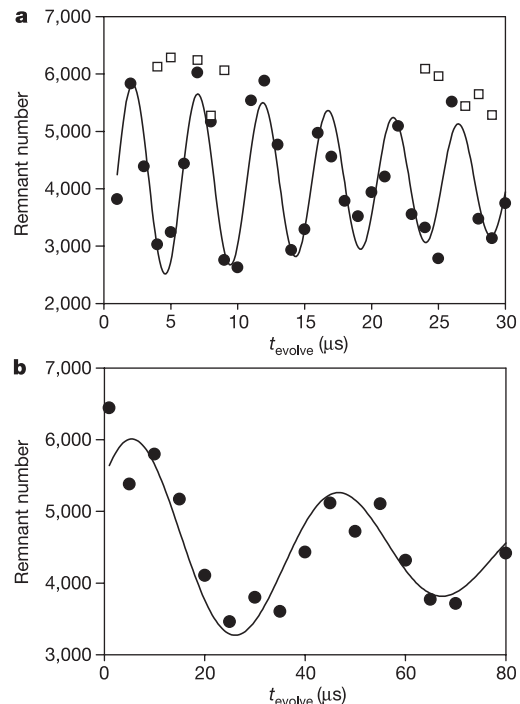


Figure 4 N_{remnant} versus t_{evolve} for $n_0 = 5.4 \times 10^{13} \text{ cm}^{-3}$. **a**, $B_{\text{evolve}} = 159.69(4)$ G ($a_{\text{evolve}} = 580a_0$). The data fitted to a damped sine wave with a frequency of $207(2)$ kHz and a decay time of $46(21) \mu\text{s}$. The open squares near $N_{\text{remnant}} = 6,000$ indicate the number remaining versus time after only pulse 1 and t_{evolve} at 159.69 G. The error estimates in parentheses are the standard errors in the units of the last decimal place. **b**, $B_{\text{evolve}} = 157.60(4)$ G ($a_{\text{evolve}} = 1,400a_0$). We note the increase in time scale from **a**. These data fitted to an oscillation frequency of $23.9(12)$ kHz and a decay time of $82(38) \mu\text{s}$.

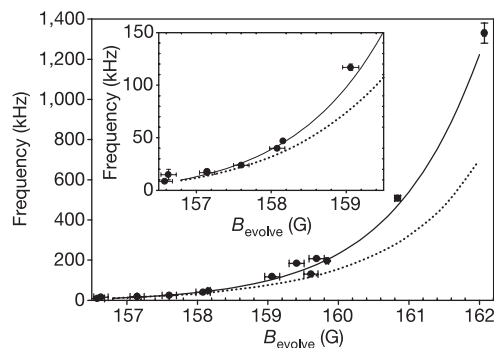


Figure 5 Oscillation frequency versus magnetic field. The points are the measured frequencies. The solid line is the energy difference between the atom–atom threshold and the bound molecular state found by S.J.J.M.F. Kokkelmans with a coupled-channels scattering calculation. The dotted line is a plot of ϵ/h . The inset is an expanded view of the lower-frequency data. The maximum frequency that we could measure was limited only by timing jitter and finite resolution in the experiment. The magnetic-field measurements for the points with the smallest horizontal error bars were performed on the same days as the corresponding frequency measurements. The error bars for the points with larger field uncertainties were inflated by 100 mG to account for estimated day-to-day field drifts.

39(4)% of the atoms were missing. Experiments with longer pulses 1 and 2 also had a higher fraction of missing atoms. For example, when we used 50 μ s pulses with $n_0 = 5.4 \times 10^{13} \text{ cm}^{-3}$, 56(3)% were missing.

We have carried out double-pulse measurements with a variety of widths and amplitudes for pulses 1 and 2 and a variety of different densities and initial magnetic fields. Although the oscillation frequency was unchanged, the phase, contrast and damping of the oscillations did vary. The contrast was very sensitive to the pulse length, and was lower for longer pulses that created more missing atoms. Defining the contrast as the oscillation amplitude divided by the time averaged number of remnant atoms detected, we observed an optimum contrast of 0.42(7) for 15 μ s pulses to 156.6(1) G. A single such pulse removed about half of the atoms from the BEC. When the pulse length was comparable to τ_{decay} , the contrast was reduced by about a factor of two. Under those conditions, three-quarters of the atoms were lost after pulse 1. The phase was shifted, but we did not observe a change in the contrast when we varied the levels of pulses 1 and 2 from $B = 156.6(1) \text{ G}$ (2,400 a_0) to $B = 155.1(1) \text{ G}$ (24,000 a_0). The contrast did depend on the intermediate level, however, and was reduced for B_{evolve} values closest to the resonance, for which the magnetic-field jumps between B_{evolve} and pulses 1 and 2 were shortest.

We also looked for a temperature dependence of both the damping and the frequency at $\sim 200 \text{ kHz}$ and did not see any. The high-temperature data were much noisier than the data for pure condensates, owing to unexplained enhanced noise in the number of thermal atoms after the magnetic-field pulse, but when the initial thermal fraction was increased from $<5\%$ to 30%, the data still showed oscillations with frequency, amplitude and damping consistent with what was observed with low-temperature data.

Our interpretation of our observations is that the first magnetic-field pulse provides a sufficiently rapid perturbation to result in nonadiabatic mixing between atomic and molecular states. The superposition then evolves according to the energy difference between the states, which is determined by the magnetic field during the free evolution stage, B_{evolve} . The second pulse mixes atom and molecule states again, such that the final state of the system depends on the relative phase of atomic and molecular fields at the time of the second pulse. This is somewhat analogous to Ramsey's method of separated oscillating fields³⁰. Over a very limited range of pulse amplitudes (a near 1,700 a_0), we could also observe Rabi-like oscillations with a single pulse towards the Feshbach resonance. The narrow observational window results

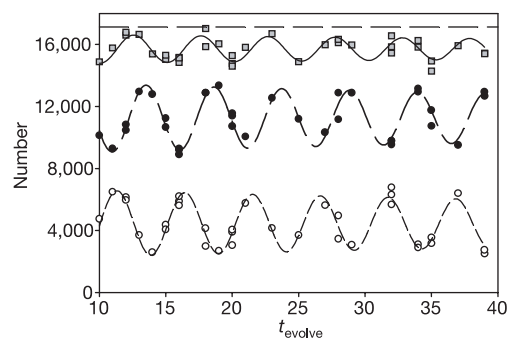


Figure 6 Number versus t_{evolve} for $n_0 = 1.1 \times 10^{13} \text{ cm}^{-3}$. From bottom to top, the data are plots of N_{burst} (open circles), N_{remnant} (filled circles), and the total number of observed atoms (grey squares). Each data set was fitted to a damped sine wave resulting in the displayed fits. $N_{\text{init}} = 17,100$ is indicated by the flat dashed line. $B_{\text{evolve}} = 159.84(2) \text{ G}$ and the remnant data fitted to an oscillation frequency of 196(1) kHz and $\tau_{\text{decay}} = 91(33) \mu\text{s}$. Not all of the data used to determine τ_{decay} are shown. To produce condensates with lower density for these measurements, the initial magnetic field before the fast-pulse sequence was 162.2(1) G and the amplitudes for pulses 1 and 2 were reduced to $\sim 7 \text{ G}$.

from the conflicting needs for both strong coupling and a condensate loss time²³ that is long compared to a Rabi oscillation period.

After pulse 2, a fraction of the coherent molecular component is converted into the energetic but still spin-polarized burst atoms through a yet to be determined process. Another mystery is the missing atoms. Are they molecules that are not converted back into atoms and are not detected in the burst or the remnant signals? If so, why do we not see them as atoms after the field is turned off and the corresponding molecular state is no longer bound? Why are there fewer missing atoms for lower-density condensates and quicker pulses towards the Feshbach resonance? What is the actual conversion efficiency from atoms to molecules and how could we maximize it? Very near the Feshbach resonance, the molecular state has a magnetic moment nearly the same as that of the free atoms, and hence will remain magnetically trapped. Another remaining question concerns the nature of the molecules. Could they be considered a molecular BEC? Clearly, there is much to be learned about this curious system. □

Received 22 April; accepted 30 April 2002.

- Wynar, R. H., Freeland, R. S., Han, D. J., Ryu, C. & Heinzen, D. J. Molecules in a Bose-Einstein condensate. *Science* **287**, 1016–1019 (2000).
- McKenzie, C. et al. Photoassociation of sodium in a Bose-Einstein condensate. *Phys. Rev. Lett.* **88**, 120403-1–120403-4 (2001).
- Anglin, J. R. & Vardi, A. Dynamics of a two-mode Bose-Einstein condensate beyond mean-field theory. *Phys. Rev. A* **64**, 013605-1–013605-9 (2001).
- Cusack, B. J., Alexander, T. J., Ostrovskaya, E. A. & Kivshar, Y. S. Existence and stability of coupled atomic-molecular Bose-Einstein condensates. *Phys. Rev. A* **65**, 013609-1–013609-4 (2001).
- Calsamiglia, J., Mackie, M. & Suominen, K. Superposition of macroscopic numbers of atoms and molecules. *Phys. Rev. Lett.* **87**, 160403-1–160403-4 (2001).
- Drummond, P. D., Kheruntsyan, K. V., Heinzen, D. J. & Wynar, R. H. Stimulated Raman adiabatic passage from an atomic to a molecular Bose-Einstein condensate. Preprint cond-mat/0110578 at (<http://xxx.lanl.gov>) (2002).
- Heinzen, D. J., Wynar, R., Drummond, P. D. & Kheruntsyan, K. V. Superchemistry: dynamics of coupled atomic and molecular Bose-Einstein condensates. *Phys. Rev. Lett.* **84**, 5029–5033 (2000).
- Tiesinga, E., Moerdijk, A., Verhaar, B. J. & Stoof, H. T. C. Conditions for Bose-Einstein condensation in magnetically trapped atomic cesium. *Phys. Rev. A* **46**, R1167–R1170 (1992).
- Tiesinga, E., Verhaar, B. J. & Stoof, H. T. C. Threshold and resonance phenomena in ultracold ground-state collisions. *Phys. Rev. A* **47**, 4114–4122 (1993).
- Moerdijk, A. J., Verhaar, B. J. & Axelson, A. Resonances in ultracold collisions of ^6Li , ^7Li , and ^{23}Na . *Phys. Rev. A* **51**, 4852–4861 (1995).
- van Abeelen, F. A. & Verhaar, B. J. Time-dependent Feshbach resonance scattering and anomalous decay of a Na Bose-Einstein condensate. *Phys. Rev. Lett.* **83**, 1550–1553 (1999).
- Mies, F. H., Tiesinga, E. & Julienne, P. S. Manipulation of Feshbach resonances in ultracold atomic collisions using time-dependent magnetic fields. *Phys. Rev. A* **61**, 022721-1–022721-17 (2000).
- van Abeelen, F. A., Heinzen, D. J. & Verhaar, B. J. Photoassociation as a probe of Feshbach resonances in cold-atom scattering. *Phys. Rev. A* **57**, R4102–R4105 (1998).
- Timmermans, E., Tommasini, P., Hussein, M. & Kerman, A. Feshbach resonances in atomic Bose-Einstein condensates. *Phys. Rep.* **315**, 199–230 (1999).

15. Timmermans, E., Tommasini, P., Côté, R., Hussein, M. & Kerman, A. Rarefied liquid properties of hybrid atomic-molecular Bose-Einstein condensates. *Phys. Rev. Lett.* **83**, 2691–2691 (1999).
16. Drummond, P. D., Kheruntsyan, K. V. & He, H. Coherent molecular solitons in Bose-Einstein condensates. *Phys. Rev. Lett.* **81**, 3055–3058 (1998).
17. Holland, M., Park, J. & Walser, R. Formation of pairing fields in resonantly coupled atomic and molecular Bose-Einstein condensates. *Phys. Rev. Lett.* **86**, 1915–1918 (2001).
18. Goral, K., Gajda, M. & Rzążewski, K. Multimode dynamics of a coupled ultracold atomic-molecular system. *Phys. Rev. Lett.* **86**, 1397–1401 (2001).
19. Vardi, A., Yurovsky, V. A. & Anglin, J. R. Quantum effects on the dynamics of a two-mode atom-molecule Bose-Einstein condensate. *Phys. Rev. A* **64**, 063611–1–063611-5 (2001).
20. Stenger, J. *et al.* Strongly enhanced inelastic collisions in a Bose-Einstein condensate near Feshbach resonances. *Phys. Rev. Lett.* **82**, 2422–2425 (1999).
21. Cornish, S. L., Claussen, N. R., Roberts, J. L., Cornell, E. A. & Wieman, C. E. Stable ^{85}Rb Bose-Einstein condensates with widely tunable interactions. *Phys. Rev. Lett.* **85**, 1795–1798 (2000).
22. Claussen, N. R., Cornish, S. L., Roberts, J. L., Cornell, E. A. & Wieman, C. E. in *Atomic Physics 17* (eds Arimondo, E., DeNatale, P. & Inguscio, M.) 325–336 (American Institute of Physics, New York, 2001).
23. Claussen, N. R., Donley, E. A., Thompson, S. T. & Wieman, C. E. Microscopic dynamics in a strongly interacting Bose-Einstein condensate. *Phys. Rev. Lett.* (submitted); preprint cond-mat/0201400 at (<http://xxx.lanl.gov>) (2002).
24. Roberts, J. L. *et al.* Controlled collapse of a Bose-Einstein condensate. *Phys. Rev. Lett.* **86**, 4211–4214 (2001).
25. Pérez-García, V. M., Michinel, H., Cirac, J. I., Lewenstein, M. & Zoller, P. Dynamics of Bose-Einstein condensates: variational solutions of the Gross-Pitaevskii equations. *Phys. Rev. A* **56**, 1424–1432 (1997).
26. Donley, E. A. *et al.* Dynamics of collapsing and exploding Bose-Einstein condensates. *Nature* **412**, 295–299 (2001).
27. Sakurai, J. J. *Modern Quantum Mechanics* 410–416 (Addison-Wesley, Reading, Massachusetts, 1994).
28. Roberts, J. L. *et al.* Improved characterization of elastic scattering near a Feshbach resonance in ^{85}Rb . *Phys. Rev. A* **64**, 024702–1–024702-3 (2001).
29. van Kempen, E. G. M., Kokkelmans, S. J. J. M. E., Heinzen, D. J. & Verhaar, B. J. Interisotope determination of ultracold rubidium interactions from three high-precision experiments. *Phys. Rev. Lett.* **88**, 093201–1–093201-4 (2002).
30. Ramsey, N. F. A molecular beam resonance method with separated oscillating fields. *Phys. Rev.* **78**, 695–699 (1950).

Acknowledgements

We acknowledge contributions from E. A. Cornell and the JILA quantum gas collaboration. We are grateful to C. H. Greene and S. J. J. M. F. Kokkelmans for providing the coupled-channels scattering calculations presented in Fig. 5 and to L. Pitaevskii for numerous discussions. S.T.T. acknowledges the support of an ARO-MURI Fellowship. This work was also supported by ONR and NSF.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.R.C. (e-mail: nclausse@jilaui.colorado.edu).

Quantum control of energy flow in light harvesting

Jennifer L. Herek^{†‡}, Wendel Wohlleben^{†‡}, Richard J. Cogdell[§], Dirk Zeidler[‡] & Marcus Motzkus[‡]

^{*} Chemical Physics, Lund University, 22100 Lund, Sweden and Department of Biophysics, Vrije Universiteit, 1081 HV Amsterdam, The Netherlands

[‡] Max Planck Institut für Quantenoptik, 85748 Garching, Germany

[§] Division of Biochemistry and Molecular Biology, IBLS, University of Glasgow, G12 8QQ, UK

[†] These authors contributed equally to this work

Coherent light sources have been widely used in control schemes that exploit quantum interference effects to direct the outcome of photochemical processes. The adaptive shaping of laser pulses is a particularly powerful tool in this context: experimental output as feedback in an iterative learning loop refines the applied laser field to render it best suited to constraints set by the experimenter^{1,2}. This approach has been experimentally implemented to control a variety of processes^{3–9}, but the extent to which

coherent excitation can also be used to direct the dynamics of complex molecular systems in a condensed-phase environment remains unclear. Here we report feedback-optimized coherent control over the energy-flow pathways in the light-harvesting antenna complex LH2 from *Rhodospseudomonas acidophila*, a photosynthetic purple bacterium. We show that phases imprinted by the light field mediate the branching ratio of energy transfer between intra- and intermolecular channels in the complex's donor-acceptor system. This result illustrates that molecular complexity need not prevent coherent control, which can thus be extended to probe and affect biological functions.

The function of an antenna complex is to harvest sunlight and, through a series of energy transfer steps both within and between complexes, to make the energy available to the reaction centre, where it is used to fuel the primary dark reactions of photosynthesis^{10–12}. The peripheral antenna LH2 from *R. acidophila* contains two pigment species, carotenoid and bacteriochlorophyll *a* (bchl), arranged in concentric rings within a protein matrix¹³ (Fig. 1a). The bchl molecules are found in two distinct binding sites and are labelled according to their absorption maxima as B800 and B850. The carotenoid pigments (rhodopin glucoside) absorb at higher energy, in the region 450–540 nm (Fig. 1c). In this particular species, only 50% of the light harvested by carotenoid is transported to the bchl acceptor states; the remaining energy is dissipated through intramolecular internal conversion processes^{14,15}. Here we explore the competition between intra- and intermolecular channels with shaped excitation light.

The ultrafast dynamics in the LH2 system are well characterized^{15–17} (Fig. 1b). Deactivation of the initially excited carotenoid singlet *S*₂ state occurs in less than 100 fs. The internal conversion (IC) channel deposits energy to the lower-lying *S*₁ state, which undergoes further non-radiative decay to the ground state in about 4 ps. The energy transfer (ET) processes involve both B800 and B850 bchls, but the final acceptor state is the B850 *Q*_Y level, the lowest singlet excited state in the complex.

To find out whether the fate of excitation energy harvested by carotenoid can be influenced by specially designed excitation pulses, we construct a learning loop^{1,2} consisting of a pulse shaper, the LH2 sample and an evolutionary algorithm strategy. We excite the carotenoid to the *S*₂ state with shaped femtosecond pulses¹⁸ from a non-collinear optical parametric amplifier (nc-OPA)¹⁹ and monitor the ensuing energy flow (Fig. 2) by probing the transient absorption kinetics of carotenoid (*S*₁–*S*_n transition, 580 nm) and bchl (B850 bleach/stimulated emission, 880 nm). These spectrally separated, discrete signals of the IC and ET channels (Fig. 2, inset) provide feedback for a learning loop evolutionary algorithm²⁰ that optimizes the excitation pulse shape for a given target. Shaping the temporal envelope and phase of a femtosecond pulse is accomplished indirectly in the frequency domain. A 4-f-zero-dispersion compressor^{18,21} creates a Fourier plane where a liquid-crystal mask is placed that modulates the spectral phase ($\phi(\lambda)$) and amplitude ($A(\lambda)$).

We begin with conventional 'blind' optimizations, where each frequency component of the pulse spectrum is initially set to have random phase from 0 to 2π and amplitude from 4 to 100%. The energy of the unshaped transform-limited pulse is slightly beyond the onset of saturation of the *S*₂ transition; this makes possible multi-photon interactions, that is, fast population cycling between electronic states. Linear interpolation across every four pixels creates smooth variations in phase and amplitude and results in 32 parameters each for phase and amplitude modulation. The target for the evolutionary algorithm is to maximize the ratio of the IC/ET signals. The optimal pulse is 30% lower in energy and yet the total excitation rate is higher. The optimization achieves an approximately 35% enhancement of the feedback signal (Fig. 3). We carefully checked that the spot overlap on the sample did not move as a result of any spatio-temporal shaping side effects²². In

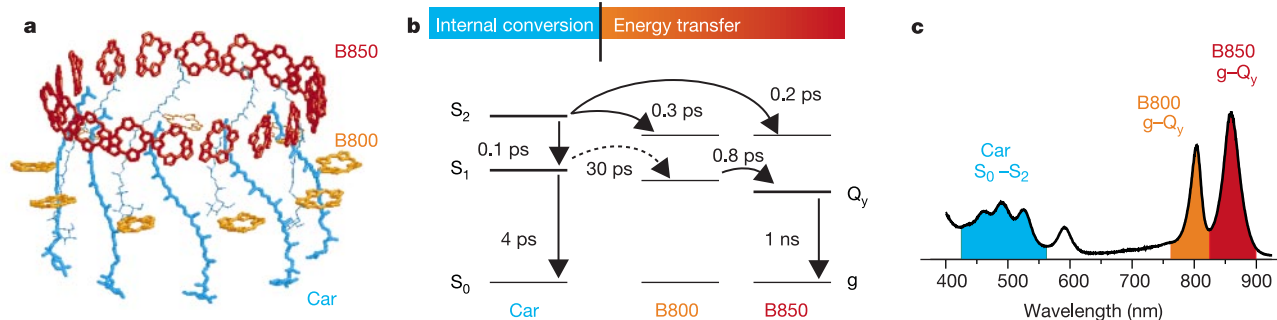


Figure 1 Properties of the LH2 antenna complex of *R. acidophila*. **a**, Geometrical arrangement of the carotenoid ('Car') and bchl molecules (data taken from the Protein Data Bank; identification code 1kzu). The carotenoid molecules are depicted in blue, B800 bchls are in orange, and B850 bchls are in dark red. **b**, Schematic energy level diagram,

indicating energy transfer (ET) pathways and time constants from the carotenoid S₂ state. There is negligibly small cross-talk, that is, transfer from the internal conversion (IC) state S₁ to bchl is around 5%. The energy funnel precludes bchl → carotenoid back transfer¹⁵. **c**, Steady state absorption spectrum.

addition, we recorded the full transient absorption spectrum (by dispersing the probe light through a monochromator), and verified that the shaped pulse affects the amplitudes of the carotenoid and bchl bands uniformly. The optimum solution (Fig. 3, inset) found by the learning algorithm is spectrally complex but has a clear temporal structure of equispaced subpulses.

Optimal control theory calculations on model systems indicate that complex pulse shapes can often be simplified to a physically intuitive shape with comparable efficiency^{23,24}. As the 'blind' optimizations described above consistently found solutions based on multi-pulse excitations²⁵, we now limit the pulse shaper by use of a parameterization²⁶ to the restricted class of pulse sequences (equispaced phase coherent subpulses). A sinusoidal phase function, $\phi(n) = a \sin(b\lambda(n) + c)$ where n is the pixel index (1–128), is applied to the pulse shaper, with only three parameters for optimization. The inter-pulse spacing of the comb is determined by the frequency parameter b , and parameter a determines the modulation

depth. The phase parameter c causes a well-defined shift of the phase pattern on the mask, leading to a change of the relative temporal phase between consecutive subpulses²⁷. In contrast to recent work²⁸, the absolute phase is of no importance here. The target of the evolutionary algorithm is again to optimize for the maximum IC/ET ratio. The algorithm reaches convergence on all three parameters and the optimal shape consists reproducibly of seven major subpulses separated by 250 ± 10 fs, where each subpulse has a duration close to the transform limit of about 30 fs. Both channels are enhanced in comparison to a single unshaped transform-limited pulse (Fig. 4), because the multi-pulse solution minimizes saturation effects. We now shift the phase pattern of the subpulse comb by adding π to the optimized value for parameter c . Mechanisms that do not involve coherence in molecular excitations will not be sensitive to such changes. However, the kinetics from the phase-shifted subpulse comb are markedly different from the optimized solution (Fig. 4).

In testing the optimized pulse envelope and its phase-shifted companion at different energies, we find that the coherent effect on the carotenoid stays constant from 15 to 250 nJ. The kinetics of the bchl B850 signal show some annihilation²⁹, making energy-flow determination at high excitation energy difficult (owing to the close coupling in the B850 ring, energy transfer from several carotenoids into the single excitonic state Q_y annihilates within about 2 ps to one surviving exciton). To quantify the effect on the energy flow-

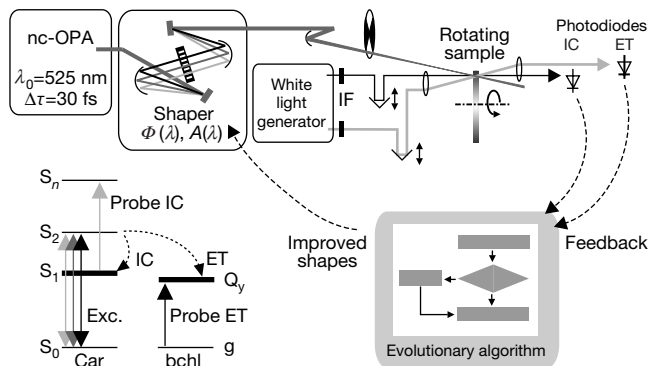


Figure 2 Optical set-up. The femtosecond pulses of a non-collinear optical parametric amplifier¹⁹ are shaped with a 2×128 pixel liquid crystal mask in the Fourier plane of a 4f-set-up¹⁸. The excitation light has a central wavelength (λ_0) of 525 nm and the duration of a transform-limited pulse is 30 fs FWHM (full-width at half-maximum). The maximum power is 250 nJ on a ~ 100 - μ m spot, corresponding to a flux of 8×10^{15} photons cm^{-2} . Two white light continua (with independent delay lines) are used for probing the carotenoid S₁-S_n excited state absorption (580 nm) and B850 bchl Q_y bleach/stimulated emission (880 nm) (inset). The spectral windows of the probe beams are set by 10 nm interference filters (IF) and the spot sizes are around 40 μ m. The transmitted light of both channels is collected on photodiodes and the change in optical density due to the action of the excitation is determined by: $\Delta\text{OD} = -\log(\text{PD}_{\text{exc}}/\text{PD}_{\text{not exc}})$ where PD_{exc} (PD_{not exc}) is the collected light intensity at the photodiode with (without) excitation. An evolutionary algorithm²⁰ is used to optimize excitation pulse shapes according to a target, for example: maximizing the IC/ET ratio. Isolated LH2 complexes of *R. acidophila* (in 50 mM Tris/HCl (pH 8) with 0.3% LDAO) are prepared with OD_{525 nm} = 0.3 and held in a 2-mm path length rotating cell at room temperature (296 K). No signs of sample degradation were observed throughout the course of the measurements.

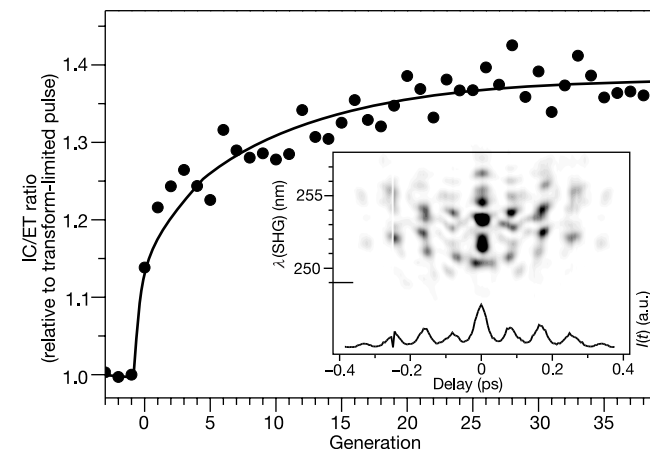


Figure 3 Optimization results. Convergence curve of the 'blind' optimization (target: to maximize IC/ET) with free phase and amplitude shaping (64 independent parameters) in the Fourier plane. The data are normalized to the ratio obtained with unshaped transform-limited pulses, corresponding to the first three points. The inset shows the optimal pulse autocorrelation ($I(t)$) with its wavelength resolution, that is, the second harmonic generation (SHG) frequency resolved optical gating (FROG) trace³⁰.

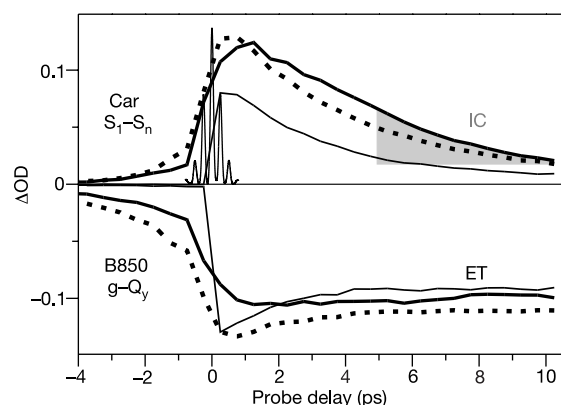


Figure 4 Transient absorption kinetics. Each curve is averaged over 15 transients with 100 shots per point. Feedback signals of the carotenoid S_1 and B850 Q_y states upon excitation with a transform-limited pulse (thin solid line), optimized subpulse comb (shown) for maximal IC/ET ratio (thick solid line) and π phase-shifted optimal pulse (dashed thick line). All pulses have equal energy (250 nJ) and spectrum. The phase-shifted pulse results in more energy flow into ET and simultaneously less into IC. For IC in Fig. 5 the carotenoid signal is integrated from 5 to 10 ps (shaded region), long after the excitation pulse, subtracting the asymptotic offset from B850 excited state absorption. For ET we take the asymptotic B850 signal. ΔOD , change in optical density.

partitioning ratio, we attenuate to 25 nJ, far below saturation and annihilation onset and compare optimized excitation shapes to phase shifted. We stress that changing the phase parameter has no measurable bearing on pulse envelope or energy. As shown in Fig. 5, the IC/ET ratio oscillates by about 30% and thus clearly confirms achievement of coherent control: relative changes of the optical carrier phase under an optimized envelope are indeed decisive for the control effect.

Complex matter waves in the internal degrees of freedom can hence be used to control the flow of excitation energy in LH2. In the present experiment the optimized pulse shape features repetitive light-matter interactions, creating coherent wavepackets that interfere in a manner that enhances carotenoid IC. Given the temporal spacing of the subpulses, this interaction most probably does not occur in the initially excited S_2 manifold, but either in the S_1 or S_0 states, and may be related to a low-frequency vibration (such as a bending mode) that induces conformational changes and affects the biological function. Studies with improved temporal resolution and other shaping capabilities will further clarify the mechanistic features of this system. Those results might then also help to explain

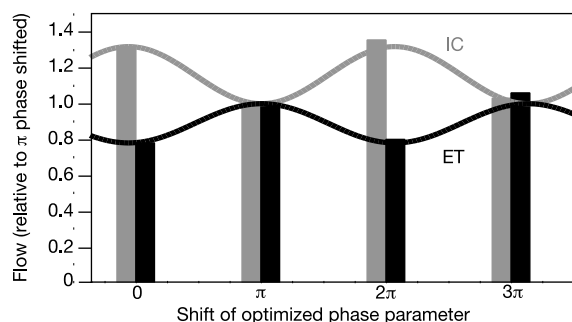


Figure 5 Evidence of coherence in the control mechanism for the light-harvesting function. First we optimize for the highest IC/ET ratio. Under this optimized temporal envelope, we then shift the phase pattern parameter c from its optimized value by 0, π , 2π , 3π . The energy-flow channels IC and ET are normalized to the lowest IC/ET ratio at π . The flow ratio oscillates between 1.32/0.78 and 1/1 by varying the phase pattern of the optimized subpulse comb envelope.

why neither the 'blind' nor the restricted sinusoidal optimization strategy yielded a pulse shape that would enhance the light-harvesting efficiency, that is, result in fewer IC losses and more energy transfer. □

Received 16 January; accepted 10 April 2002.

- Judson, R. S. & Rabitz, H. Teaching lasers to control molecules. *Phys. Rev. Lett.* **68**, 1500–1503 (1992).
- Rabitz, H., deVivie-Riedle, R., Motzkus, M. & Kompa, K.-L. Whither the future of controlling quantum phenomena? *Science* **288**, 824–828 (2000).
- Assion, A. *et al.* Control of chemical reactions by feedback-optimized pulse-shaped fs laser pulses. *Science* **282**, 919–922 (1998).
- Levis, R. J., Menkir, G. M. & Rabitz, H. Selective bond dissociation and rearrangement with optimally tailored, strong-field laser pulses. *Science* **292**, 709–713 (2001).
- Zeidler, D. *et al.* Optimal control of ground-state dynamics in polymers. *J. Chem. Phys.* **116**, 5231–5235 (2002).
- Bardeen, C. J. *et al.* Feedback quantum control of molecular electronic population transfer. *Chem. Phys. Lett.* **280**, 151–158 (1997).
- Brixner, T., Damrauer, N. H., Niklaus, P. & Gerber, G. Photosynthetic adaptive femtosecond quantum control in the liquid phase. *Nature* **414**, 57–60 (2001).
- Weinacht, T. C., Ahn, J. & Bucksbaum, P. H. Controlling the shape of a quantum wavefunction. *Nature* **397**, 233–235 (1999).
- Bartels, R. *et al.* Shaped-pulse optimization of coherent emission of high-harmonic soft X-rays. *Nature* **406**, 164–166 (2000).
- Pullerits, T. & Sundström, V. Photosynthetic light-harvesting pigment-protein complexes: Toward understanding how and why. *Acc. Chem. Res.* **29**, 381–389 (1996).
- Zinth, W., Airl, T. & Wachtveit, J. The primary processes of bacterial photosynthesis—ultrafast reactions for the optimum use of light energy. *Ber. Bunsenges. Phys. Chem.* **100**, 1962–1966 (1996).
- Ritz, T., Park, S. & Schulten, K. Kinetics of excitation migration and trapping in the photosynthetic unit of purple bacteria. *J. Phys. Chem. B* **105**, 8259–8267 (2001).
- Mc Dermott, G. *et al.* Crystal structure of an integral membrane light-harvesting complex from photosynthetic bacteria. *Nature* **374**, 517–521 (1995).
- Frank, H. A. *et al.* On the photophysics and photochemical properties of carotenoids and their role as light-harvesting pigments in photosynthesis. *Pure Appl. Chem.* **69**, 2117–2124 (1997).
- Macpherson, A., Arellano, J. B., Fraser, N. J., Cogdell, R. J. & Gillbro, T. Efficient energy transfer from the carotenoid S_2 state in a photosynthetic light-harvesting complex. *Biophys. J.* **80**, 923–930 (2001).
- Sundström, V., Pullerits, T. & van Grondelle, R. Photosynthetic light harvesting: reconciling dynamics and structure of purple bacterial LH2 reveals function of photosynthetic unit. *J. Phys. Chem. B* **103**, 2327–2346 (1999).
- Krueger, B. P., Scholes, G. D., Jimenez, R. & Fleming, G. R. Electronic excitation energy transfer from carotenoid to bacteriochlorophyll in the purple bacterium *Rhodospseudomonas acidophila*. *J. Phys. Chem. B* **102**, 2284–2292 (1998).
- Zeidler, D., Hornung, T., Proch, D. & Motzkus, M. Adaptive compression of pulses from a non-collinear OPA to <16 fs by feedback controlled pulse shaping. *Appl. Phys. B* **70**(Suppl.), S125 (2000).
- Wilhelm, T., Piel, J. & Riedle, E. Sub-20-fs pulses tunable across the visible from a blue-pumped single-pass noncollinear parametric converter. *Opt. Lett.* **22**, 1494–1496 (1997).
- Zeidler, D., Frey, S., Kompa, K.-L. & Motzkus, M. Evolutionary algorithms and their application to optimal control studies. *Phys. Rev. A* **64**, 023420-1–023420-13 (2000).
- Weiner, A. M. Femtosecond pulse shaping using spatial light modulators. *Rev. Sci. Instrum.* **71**, 1929–1960 (2000).
- Wefers, M. M. & Nelson, K. A. Space-time profiles of shaped ultrafast optical waveforms. *IEEE J. Quant. Electr.* **32**, 161–172 (1996).
- Amstrup, B., Carson, R. J., Matro, A. & Rice, S. A. The use of pulse shaping to control the photodissociation of a diatomic molecule: Preventing the best from being the enemy of the good. *J. Phys. Chem.* **95**, 8019–8027 (1991).
- Hornung, T., Motzkus, M. & de Vivie-Riedle, R. Teaching OCT to distil robust pulses even under experimental constraints. *Phys. Rev. A* **65**, 021403(R)–021403(R)-4 (2002).
- Warren, W. S. & Zewail, A. H. Multiple phase-coherent laser pulses in optical spectroscopy. I. The technique and experimental applications. *J. Chem. Phys.* **78**, 2279–2297 (1983).
- Hornung, T., Meier, R. & Motzkus, M. Optimal control of molecular states in a learning loop with parameterization in frequency and time domain. *Chem. Phys. Lett.* **326**, 445–453 (2000).
- Meshulach, D. & Silberberg, Y. Coherent quantum control of two-photon transitions by a femtosecond laser pulse. *Nature* **396**, 239–242 (1998).
- Paulus, G. G. *et al.* Absolute-phase phenomena in photoionization with few-cycle laser pulses. *Nature* **414**, 182–184 (2001).
- Trinkunas, G., Herek, J. L., Polivka, T., Sundström, V. & Pullerits, T. Exciton delocalization probed by excitation annihilation in the light-harvesting antenna LH2. *Phys. Rev. Lett.* **86**, 4167–4170 (2001).
- Trebino, R. *et al.* Measuring ultrashort laser pulses in the time-frequency domain using frequency-resolved optical gating. *Rev. Sci. Instrum.* **68**, 3277–3295 (1997).

Acknowledgements

We acknowledge support and encouragement from K.-L. Kompa and V. Sundström. We thank A. Gardiner for help with sample preparation, T. Hornung for discussions and M. Mathys for help with experiments.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.M. (e-mail: mcm@mpq.mpg.de).

An inverted continental Moho and serpentinization of the forearc mantle

M. G. Bostock^{*}, R. D. Hyndman^{†‡}, S. Rondenay^{*§} & S. M. Peacock^{||}

^{*} Department of Earth and Ocean Sciences, The University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

[†] Pacific Geoscience Centre, Geological Survey of Canada, Sidney, British Columbia V8L 4B2, Canada

[‡] School of Earth and Ocean Sciences, University of Victoria, British Columbia V8W 3P6, Canada

^{||} Department of Geological Sciences, Arizona State University, Tempe, Arizona 85287-1404, USA

Volatiles that are transported by subducting lithospheric plates to depths greater than 100 km are thought to induce partial melting in the overlying mantle wedge, resulting in arc magmatism and the addition of significant quantities of material to the overlying lithosphere¹. Asthenospheric flow and upwelling within the wedge produce increased lithospheric temperatures in this back-arc region, but the forearc mantle (in the corner of the wedge) is thought to be significantly cooler. Here we explore the structure of the mantle wedge in the southern Cascadia subduction zone using scattered teleseismic waves recorded on a dense portable array of broadband seismometers. We find very low shear-wave velocities in the cold forearc mantle indicated by the exceptional occurrence of an 'inverted' continental Moho, which reverts to normal polarity seaward of the Cascade arc. This observation provides compelling evidence for a highly hydrated and serpentinized forearc region², consistent with thermal and petrological models of the forearc mantle wedge. This serpentinized material is thought to have low strength and may therefore control the down-dip rupture limit of great thrust earthquakes, as well as the nature of large-scale flow in the mantle wedge.

In 1993–94, researchers at the Oregon State University deployed about 40 seismometers at 69 sites across central Oregon as part of an IRIS-PASSCAL experiment to investigate detailed structure of the Cascadia subduction zone³ (Fig. 1). Stations were located at intervals of 5 km, and recorded high-quality P-wave seismograms from 31 earthquakes at teleseismic (30–100°) distances. The array geometry and station density permit the application of formal, multi-channel waveform inversions of scattered waves arriving after the direct P-wave that are generated by discontinuous structure below the array^{4,5}. These scattered waves include both forward-scattered P-wave-to-S-wave conversions and back-scattered waves afforded by free-surface reflections, and are sensitive to discontinuities in shear-wave velocity and density.

In Fig. 2a we display an image of S-wave velocity perturbations beneath central Oregon as determined from the simultaneous inversion of scattered waves from the 31 earthquakes. The perturbations are defined with respect to a smoothly varying, one-dimensional reference model. The limited frequency content of the scattered wave data results in a bandpass-filtered approximation to the true perturbation structure. Planar discontinuities are therefore defined by rapid transitions from high/low to low/high velocity perturbations. Two prominent structures dominate the image in Fig. 2a. The continental Moho is evident as a boundary near 36 km depth across the eastern portions (east of –122.3° longitude) of the section, where it separates low-velocity (shown red) continental crust from high-velocity (shown blue) mantle. The Moho persists

seaward to approximately 40 km west of the arc, beyond which it apparently disappears. At the western edge of the profile we note an approximately 10-km-thick, low-velocity layer centred at 25 km depth, which dips shallowly at ~10°. As in an earlier forward modelling study employing the same data set³, this layer is associated with the oceanic crust of the subducting Juan de Fuca plate. By 45 km depth, the oceanic Moho (lower boundary of oceanic crust) dips more steeply (~30°) and exhibits a reduced velocity contrast. It can be traced to a depth of ~90 km towards the eastern end of the profile. The upper boundary of subducted Juan de Fuca crust appears to possess a landward evolution that is more complex than that of the underlying Moho. At 30 km depth, the upper-crustal boundary apparently levels off and adopts a horizontal trajectory with diminishing velocity contrast to the east. The relation of this subhorizontal boundary with the continental Moho further landward, or with the dipping oceanic Moho below, is unclear. However, there is an absence of strongly discontinuous structure in the intervening volume.

To understand the nature of these variations in velocity structure, we need to consider the water budget of the downgoing oceanic crust, the effects of hydration on mantle peridotite, and temperatures across the subduction complex. It is generally understood that significant quantities of aqueous fluid are expelled from the downgoing crust and sediments as they encounter increasing pressure and temperature⁶. Much of this fluid is driven off at shallow levels, but estimates of fluid fluxes based on core and downhole measurements⁷ suggest that, over 10–20 Myr, sufficient H₂O is released at subcrustal levels to hydrate the entire forearc mantle. Hydration of depleted peridotite (the dominant mantle rock type comprising olivine, orthopyroxene and lesser amounts of clinopyroxene and Cr-spinel) will stabilize a variety of hydrous minerals, in particular, serpentine⁸. The stable mineral assemblage will depend on temperature, pressure and bulk composition, especially SiO₂ content. In addition to serpentine minerals, hydrated peridotite may contain additional hydrous minerals, such as amphibole, brucite, chlorite and talc. Antigorite is the main serpentine mineral in ultramafic rocks metamorphosed under moderate temperatures, and is stable to temperatures of 620–720 °C at depths between 30 and 150 km (ref. 9).

In addition to supplying the fluids that hydrate the overlying

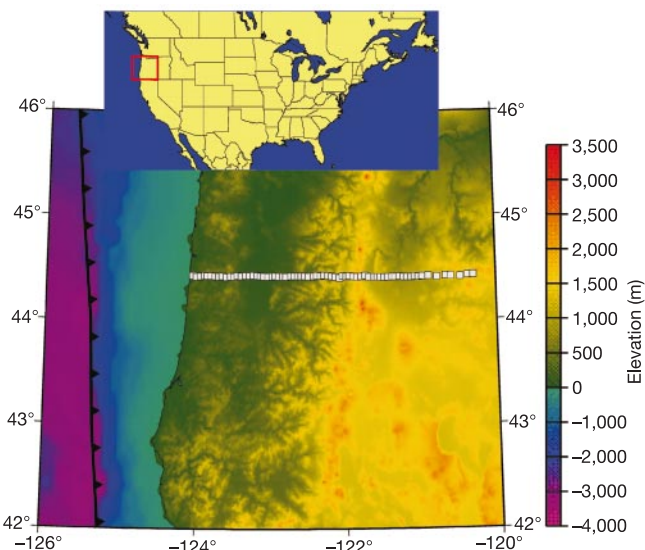


Figure 1 Relief map of central Cascadia subduction zone. Locations of broadband, three-component seismic stations that recorded the data employed in this study are shown as white squares. Note that the array is oriented approximately perpendicular to the strike of major topographic features associated with the subduction zone.

[§] Present address: Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA

mantle wedge, the subducting plate also controls the thermal structure of the subduction-zone forearc. In particular, it significantly depresses isotherms in the mantle wedge, as evidenced by the very low surface heat flow ($30\text{--}40\text{ mW m}^{-2}$) observed in most forearcs including Cascadia^{10,11}. In Fig. 2b we plot a thermal model for central Oregon¹² corresponding to the teleseismic profile,

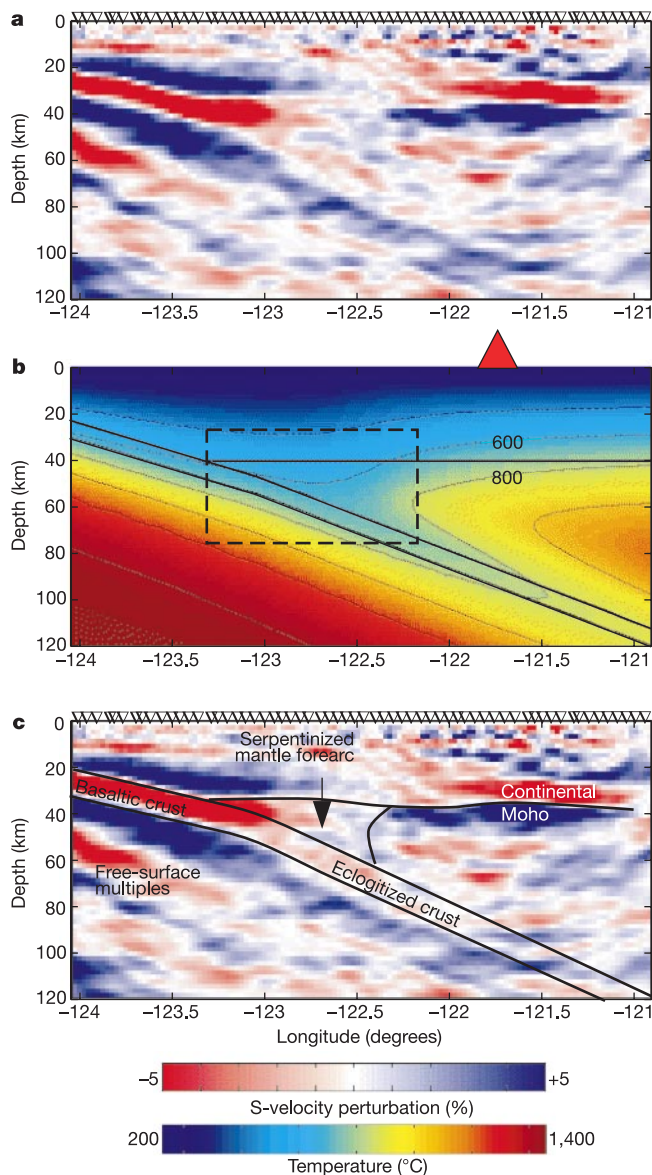


Figure 2 Comparison of scattered wave inversion results with thermal model. **a**, S-velocity perturbations below the array, recovered from the inversion of scattered waves in the P-wave coda of 31 earthquakes recorded at teleseismic distances. The image represents a bandpass-filtered version of the true perturbations to a one-dimensional, smoothly varying reference model. Discontinuities are present where steep changes in perturbation polarity occur. **b**, Thermal model of Cascadia subduction zone corresponding approximately to the profile in **a**. The cool subducting plate depresses isotherms in the forearc, rendering serpentine stable within that portion of the mantle encompassed by the dashed rectangle; solid lines indicate locations of subducting oceanic crust and continental Moho. Note temperature contour interval is 200°C . **c**, Interpretation of structure in **a**. High degrees of mantle serpentinization where the subducting oceanic crust enters the forearc mantle results in an inverted continental Moho (high-velocity crust on low-velocity mantle), which gradually reverts eastward to normal polarity by -122.3° longitude. The signature of the subducting oceanic Moho diminishes with depth as a result of progressive eclogitization below 45 km. Inverted triangles in **a** and **c** show instrument locations.

and based on heat flow and other geophysical data. In the region where oceanic crust meets the forearc mantle, temperatures are low, between 400 and 600°C , a situation that persists over a horizontal distance of $\sim 50\text{--}100$ km. However, a landward increase in heat flow from <40 to $>80\text{ mW m}^{-2}$ over ~ 20 km signals an abrupt increase in deep temperatures. In particular, Moho temperatures beneath the arc and backarc are significantly higher, above 800°C . Thus serpentine should exist in that portion of the mantle forearc contained within the dashed square in Fig. 2b, but it will not be stable beneath the arc and backarc. The degree of serpentinization in the forearc will depend upon the amount of H_2O that chemically interacts with forearc mantle, which, in turn, depends on H_2O flux from the subducting slab and the permeability structure of the slab, plate interface and mantle.

Serpentine exhibits elastic properties that are unique among commonly occurring rock types, notably low elastic wave velocities and high Poisson's ratio. The very low S-velocity of serpentine is central to the interpretation of our results. In particular, its S-velocity (v_s) is significantly lower than that of its peridotitic protolith ($\delta v_s \approx -2\text{ km s}^{-1}$) and commonly occurring lower-crustal lithologies ($\delta v_s \approx 1\text{ km s}^{-1}$)¹³. Figure 3 shows the S-wave velocity of mantle peridotite samples as a function of degree of serpentinization at a pressure of 1 GPa, which is appropriate for the base of a $\sim 35\text{-km}$ -thick continental crust as presented in ref. 14. Correction from room temperature to $400\text{--}500^\circ\text{C}$ will shift this curve downward to velocities $0.1\text{--}0.2\text{ km s}^{-1}$ lower. This information allows us to interpret the image in Fig. 2a quantitatively in terms of degree of serpentinization in the forearc mantle. As in a previous study³, we interpret the change in dip of the subducting plate by 45 km depth to indicate the onset of eclogitization of the oceanic crust, leading, eventually, to a 15% increase in density and a pronounced reduction in the seismic contrast with underlying oceanic mantle¹⁵.

Although a continuous dehydration of downgoing oceanic crust and entrained sediments is expected, the water released by eclogitization (between 1.2 and 3.3 wt%; ref. 16) is especially important for expulsion into the overlying mantle wedge, where it causes hydration and serpentinization, and significantly diminished velocities. The horizontal boundary near 32 km depth and between -122.6 and -123.3° longitude that juxtaposes high- (or neutral-) velocity material above with low-velocity material below is thus inferred to manifest the highly unusual occurrence of an 'inverted' continental Moho separating lower-crustal rocks from underlying,

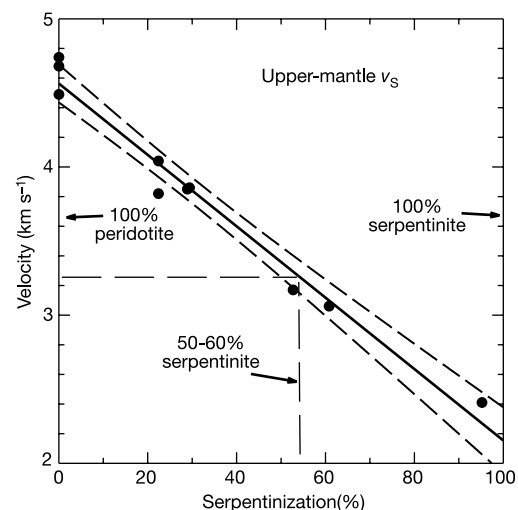


Figure 3 S-velocity of altered peridotite as a function of degree of serpentinization. Data from ref. 14. Bold line shows best-fit linear regression with $\pm 1\sigma$ error bounds. The predicted velocity contrast at the wedge corner suggests degrees of serpentinization as high as 50–60%. v_s , S-wave velocity.

serpentinized peridotite (Fig. 2c). Assuming a lower-crustal velocity of 3.6 km s^{-1} , the estimated maximum S-velocity perturbation of 10% near the wedge corner could represent a level of serpentinization as high as 50–60%. However, steadily diminishing levels are implied away from the corner as the velocity perturbation decreases through zero to normal polarity. Permeability within altered peridotite is likely to be fracture controlled, and the presence of free fluids may also serve to lower velocities. The degree of hydration diminishes eastwards, producing, in succession, reduced velocity contrast, disappearance of the aberrant Moho and, finally, restoration to a normal polarity crust–mantle boundary 40 km west of the arc.

In addition to a thermal model of Cascadia, we have produced thermal models for more typical cold continental (for example, northeast Japan) and oceanic (for example, Izu-Bonin) subduction zones. In cold subduction zones, the forearc temperatures are still lower, and in oceanic subduction zones with thin crust, the forearc mantle region is often larger. Thus although dimensions and geometry vary, all models predict the presence of a hydrated (and, hence, serpentinized) mantle wedge, suggesting that such structures are a generic feature of convergent margins. This inference is supported by recent studies in central Japan¹⁷ and northern Chile¹⁸ that employed travel times from local earthquakes to identify material with high Poisson's ratio above the subducting plate with serpentinized peridotite. Documentation of active serpentine mud volcanoes in the Mariana forearc¹⁹ provides further supporting evidence.

The ubiquitous presence of a serpentinized mantle forearc in subduction zones has at least two important implications. It has been noted previously that the down-dip rupture limit of great subduction zone earthquakes in cold subduction zones usually coincides in depth with the continental Moho²⁰. This observation can be explained through the presence of an altered mantle wedge²¹. Serpentine^{22–24}, and its companion alteration products brucite and talc⁸, are believed to exhibit stable sliding behaviour at plate velocities, thus impeding rupture into the forearc mantle. A second consequence is that large-scale flow in the mantle wedge will be modified by the presence of serpentinized forearc mantle. In particular, the positive buoyancy and weak rheology of serpentine should serve to isolate most of the hydrated forearc wedge from the mantle-wedge corner flow system, while basal portions could be dragged down by the subducting plate. □

Received 21 December 2001; accepted 15 April 2002.

- Gill, J. *Orogenic Andesites and Plate Tectonics* (Springer, New York, 1981).
- Fyfe, W. S. & McBirney, A. R. Subduction and the structure of andesitic volcanic belts. *Am. J. Sci. A* **275**, 285–297 (1975).
- Nabelek, J. L. *et al.* A high-resolution image of the Cascadia subduction zone from teleseismic converted phases recorded by a broadband seismic array. *Eos* **74**, 431 (1993).
- Bostock, M. G., Rondenay, S. & Shragge, J. Multiparameter two-dimensional inversion of scattered teleseismic body waves 1. Theory for oblique incidence. *J. Geophys. Res.* **106**, 30771–30782 (2001).
- Rondenay, S., Bostock, M. G. & Shragge, J. Multiparameter two-dimensional inversion of scattered teleseismic body waves 3. Application to the Cascadia 1993 data set. *J. Geophys. Res.* **106**, 30795–30807 (2001).
- Peacock, S. M. Large-scale hydration of the lithosphere above subducting slabs. *Chem. Geol.* **108**, 49–59 (1993).
- Fisher, A. T. Permeability within basaltic oceanic crust. *Rev. Geophys.* **36**, 143–182 (1998).
- Peacock, S. M. & Hyndman, R. D. Hydrous minerals in the mantle wedge and the maximum depth of subduction thrust earthquakes. *Geophys. Res. Lett.* **26**, 2517–2520 (1999).
- Ulmer, P. & Trommsdorff, V. Serpentine stability to mantle depths and subduction-related magmatism. *Science* **268**, 858–861 (1995).
- Blackwell, D. D. *et al.* Heat flow in the Oregon Cascade Range and its correlation with regional gravity, Curie point depths, and geology. *J. Geophys. Res.* **95**, 19475–19493 (1990).
- Lewis, T. J. *et al.* Subduction of the Juan de Fuca plate: thermal consequences. *J. Geophys. Res.* **93**, 15207–15227 (1988).
- McMahon, A. *The Thermal Structure of the Cascadia Subduction Zone: Implications for Arc Magma Generation* Thesis, Arizona State Univ., Tempe (2001).
- Christensen, N. Poisson's ratio and crustal seismology. *J. Geophys. Res.* **101**, 3139–3156 (1996).
- Christensen, N. Elasticity of ultrabasic rocks. *J. Geophys. Res.* **71**, 5921–5931 (1966).
- Ahrens, T. J. & Schubert, G. Gabbro-eclogite reaction rate and its geophysical significance. *Rev. Geophys. Space Phys.* **13**, 383–400 (1975).
- Hacker, B. R. *et al.* in *Subduction Top to Bottom* (ed. Bebout, G. E.) 105–111 (American Geophysical Union, Washington DC, 1996).
- Kamiya, S. & Kobayashi, Y. Seismological evidence for the existence of serpentinized wedge mantle. *Geophys. Res. Lett.* **27**, 819–822 (2000).
- Graeber, F. M. & Asch, G. Three-dimensional models of P wave velocity and P-to-S velocity ratio in the

- southern central Andes by simultaneous inversion of local earthquake data. *J. Geophys. Res.* **104**, 20237–20256 (1999).
- Fryer, P. Evolution of the Mariana convergent plate margin System. *Rev. Geophys.* **34**, 89–125 (1996).
- Tichelaar, B. W. & Ruff, L. J. Depth of seismic coupling along subduction zones. *J. Geophys. Res.* **98**, 2017–2037 (1993).
- Oleskevich, D. A., Hyndman, R. D. & Wang, K. The updip and downdip limits to great subduction earthquakes: thermal and structural models of Cascadia, south Alaska, SW Japan and Chile. *J. Geophys. Res.* **104**, 14965–14991 (1999).
- Escartin, J., Hirth, G. & Evans, B. Strength of slightly serpentinized peridotites: Implications for the tectonics of oceanic lithosphere. *Geology* **29**, 1023–1026 (2001).
- Raleigh, C. B. & Paterson, M. S. Experimental deformation of serpentinite and its tectonic consequences. *J. Geophys. Res.* **70**, 3965–3985 (1965).
- Reinen, L. A. Seismic and aseismic slip indicators in serpentinite gouge. *Geology* **28**, 135–138 (2000).

Acknowledgements

We thank J. Nabelek and co-workers at Oregon State University for collecting the data set, and R. Benson for facilitating access to the IRIS DMC data archive.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.G.B. (e-mail: bostock@geop.ubc.ca).

Common mammals drive the evolutionary increase of hypsodonty in the Neogene

Jukka Jernvall*† & Mikael Fortelius‡

* Developmental Biology Programme, Institute of Biotechnology, PO Box 56, FIN-00014, University of Helsinki, Helsinki, Finland

† Department of Ecology and Systematics, PO Box 65; ‡ Department of Geology, PO Box 64, FIN-00014 University of Helsinki, Finland

During the past 20 million years, herbivorous mammals of numerous lineages have evolved hypsodont, or high-crowned, cheek teeth. Hypsodonty is informative ecologically because it is well developed in mammals eating fibrous and abrasive foods that are most abundant in open and generally or seasonally dry environments^{1–5}. Here we report that in the Neogene of Europe mammals with the greatest locality coverages showed an increase in hypsodonty. We used a data set of 209 localities to measure whether large mammals occurring in many fossil localities show a similar increase in hypsodonty to mammals occurring in single or few localities. Taxonomic and morphological groupings show a low average hypsodonty in the early Miocene epoch. From the middle Miocene onwards, only the hypsodonty of commonly found mammals shows a marked increase. Therefore, in the drying Europe of the late Miocene, only increasingly hypsodont mammals may have been able to expand their share of habitats and food resources. These results suggest that the relatively small number of species known from multiple localities are palaeoecologically informative by themselves, irrespective of the rest of the known species.

Palaeoecological inference relies on information derived from individual fossil localities. Increased sampling and methods that correct for uneven sampling^{6,7} or are insensitive to sampling^{8,9} allow a more accurate, less biased description of evolutionary patterns. Large data sets including numerous localities now allow us to evaluate new aspects of community evolution, such as whether species found in many localities show similar palaeoecological

changes to those of species found in a limited number of localities. This information is lost when conventional tabulations of regional faunal lists are used for analysis, because every species is tabulated only once. Such analyses tend to over-emphasize rare taxa, whose specializations may not reflect overall trends in the biota.

We used the NOW database of Eurasian Neogene fossil mammals (<http://www.helsinki.fi/science/nov/>), from which we included European localities up to 20° E longitude. This area corresponds roughly to 'West' of previous analyses of western Eurasian Neogene mammals, an area found to represent a biogeographically and biochronologically relatively uniform region⁹; thus providing a good geographic scale for exploring the effects of environmental change on biota¹⁰. The localities from 18 to 5 million years (Myr) ago were divided into nine intervals corresponding to the MN (Mammal Neogene) units 4 to 13 of European mammal chronology^{11–13}. The number of localities in each interval ranged from 10 to 37 with a relatively constant geographic coverage (Fig. 1). We present here the results after grouping localities that are within one degree of geographic separation from each other in each time interval. This grouping makes the analysis more conservative as multiple-locality complexes (such as Oberdorf, Austria) may bias the results, although equivalent results were obtained using the ungrouped set of localities. We included all large, non-carnivorous mammals in the analysis. Three classes of tooth crown height recorded in the database were assigned the values 1, 2 and 3, respectively: low (brachydont), medium high (mesodont) and high (hypsodont). This is a relatively conservative tabulation as the difference in crown height between a hypsodont and a brachydont species is usually more than 3:1. We tabulated hypsodonty values for species, genera and cheek tooth crown types. We placed more emphasis on the genus and crown-type level tabulations because even after correcting for known synonyms (see the NOW database above), species level inferences may be biased by the high number of species found at very few localities (36 to 57 per cent of species have one locality occurrence in each MN unit). The ecomorphological crown type tabulation is based on tooth morphology alone and cuts across taxonomic groups^{14,15}.

First we examined changes in hypsodonty in taxa that are common (defined as present in more than 25 per cent of localities within an MN interval) as opposed to taxa which are rare (represented in less than 25 per cent of localities). The results show that

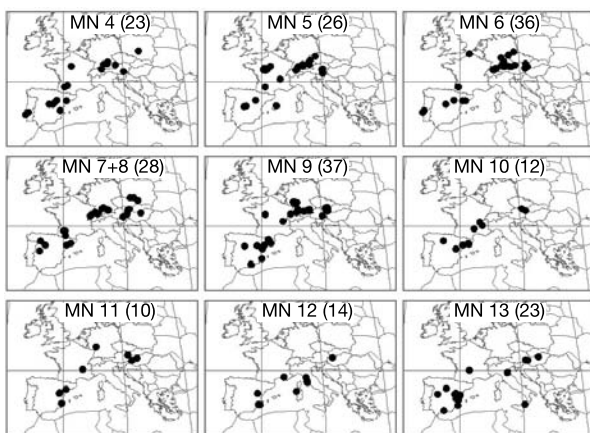


Figure 1 The nine locality intervals and number of localities (in parenthesis) analysed. The corresponding number of localities for each MN unit after grouping localities situated within one degree of geographic separation from each other are: 15, 15, 20, 22, 10, 9, 7 and 14. Grouping the short-duration MN units 10 and 11 does not change the results. The average centroids along the longitude and the latitude of the taxa or crown types found in many localities (Fig. 2) did not differ from those found in fewer localities ($P = 0.57$ to 0.76 , Mann–Whitney U -test).

the average hypsodonty increased markedly in the late Miocene in common genera (Fig. 2a). In contrast, rare genera show less change in their hypsodonty. Moreover, in the early Miocene rare genera show slightly higher hypsodonty averages than common genera (Fig. 2a). We next examined species grouped only by molar crown type. Individual crown types are sampled from more localities than genera because several taxa can share the same crown type but the pattern of results nevertheless appears similar between genera and crown types. Average hypsodonty again shows a marked increase only among common crown types in the late Miocene (Fig. 2b). This increase partly coincides with a decrease in overall sampling intensity as reflected in the number of localities known (Fig. 2c, d). Because we define frequency of occurrence in relative terms within each time interval, our analysis should be insensitive to differences in level of sampling.

As the crown type tabulation suggests, the increase in hypsodonty among common taxa is not limited to any single taxonomic group. Of special interest are the hipparionine horses, which contribute heavily to the classic increase in hypsodonty among ungulates¹. However, the increase in hypsodonty in common genera or crown types is not eliminated by removing the family Equidae or corresponding crown types from the common groups (Fig. 2a, b). Rather, hipparionine horses, many of which are found in many localities, appear as the first wave of the increase in hypsodonty followed by other ungulates later in the Neogene (Fig. 2a, b). This is evidently a result of their having evolved hypsodonty in North America before their dispersal to the Old World at the beginning of the late Miocene^{1,16}.

A significant aspect of the results is the increase in hypsodonty among the common groups beginning around 11 Myr ago. This was a time of major climatic changes, including the first development of Eurasian mid-latitude aridity¹⁷, widely believed to have been ultimately driven by the Himalayan–Tibetan uplift^{18,19}. Up to that

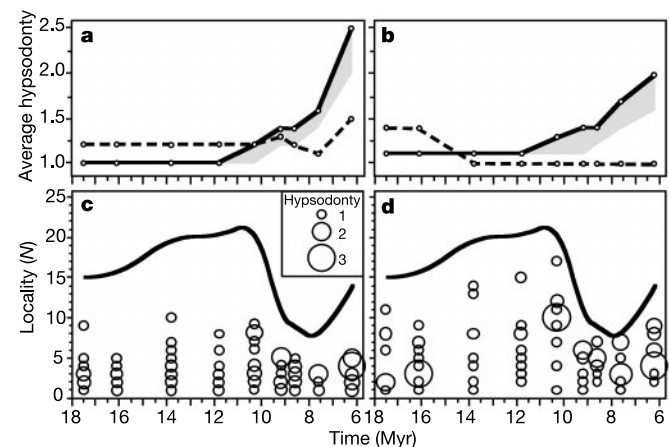


Figure 2 Increase in hypsodonty is limited to common groups (present in over 25 per cent of localities within MN interval). **a**, Average hypsodonty increases sharply in common genera (solid line) while rare taxa show very little change in hypsodonty (present in less than 25 per cent of localities, dotted line). **b**, Ecomorphological groupings of teeth based on crown types shows increase in hypsodonty among common crown types (solid line). We note how in the early Miocene, the common groups are less hypsodont than the rare groups. The increase in hypsodonty persists when hipparionine horses are removed from the common groups (in grey in **a** and **b**). **c**, The distribution of genera is well below the number (N) of localities (solid line) whereas the distribution of crown types (**d**) approaches the locality curve, representing the broader cross-taxonomic range of ecomorphological groupings. We note how the average hypsodonty of genera and crown types found in many localities increase in the late Miocene (large bubbles). We show genera in **a** and **c** because genus level can be considered a more robust measure of diversity. Species level tabulations gave equivalent results to the tabulation based on genera. See Methods for details.

point, hypsodont mammals are confined to fewer than five localities (Fig. 2c, d). Comparing further the pre- and post-11-Myr-times, average locality coverage of hypsodont species (tabulations 2 and 3) increases from 11 to 18 per cent ($P = 0.017$, MN-randomization test on medians) and for brachydont species (tabulation 1) from 13 to 16 per cent ($P = 0.054$). The locality coverage of hypsodont genera increases from 12 to 25 per cent ($P = 0.022$) and for brachydont genera from 16 to 18 per cent ($P = 0.148$). Similarly, locality coverage of hypsodont crown types increases from 8 to 52 per cent ($P = 0.028$) and for brachydont crown types from 30 to 33 per cent ($P = 0.570$). These changes are consistent with the documented decrease in overall provinciality⁹, here shown to be most conspicuous in hypsodont mammals. We propose that in the drying Europe of the late Miocene mostly mammalian lineages possessing advanced (such as the hipparionine horses) or incipient (such as many rhinoceroses and bovids) hypsodonty were able to increase their relative share of habitats and food resources. At the population level this increasing presence of hypsodont taxa may have been related to, and consistent with, life-history differences observed between modern hypsodont and brachydont ungulates, such as large group sizes found in hypsodont taxa^{20–22}. We note that while common taxa (as defined above) appear to be also abundant in relative numbers of individuals within a locality (for example, the hipparionine horses), this will require further study when locality-specific taxon abundance data become available.

The more pronounced increase in hypsodonty among common taxa is an indication that large-scale data sets provide unique insights into evolution. The most immediate implication is that common taxa could be used to retrieve and map evolutionary changes that might otherwise be swamped by noise from rare groups. Moreover, to the extent that the occurrence of fossil species in a locality may reflect past abundance²³, the use of regional or global taxon presence lists may underestimate past ecological changes at the community level. We thus predict that analysis of other large-scale data sets with primary locality–occurrence information will reveal that common taxa show generally stronger evolutionary trends than do all taxa. □

Methods

We analysed hypsodonty in species, genera and crown types of orders Artiodactyla, Perissodactyla, Primates and Proboscidea. Analyses were insensitive to details of handling genera or crown types that contained species with different hypsodonty values, and here we used averaged hypsodonty for each locality. The MN-wide hypsodonty was the average of all the localities. Four genera (*Alicornops*, *Brachypotherium*, *Plesiaceratherium* and *Tragoportax*) out of 98 genera and three crown types (L2212 (ectolophodont), L2220 (selenodont) and R2200 (bunodont)) out of 12 crown types had species with different hypsodonty values within MN units. These were also the only groups that had species with different hypsodonty values across MN units, predominantly showing an increase in hypsodonty during their temporal ranges. We assigned uncertain species (20) and genera (4) to be certain, although alternative handling did not change the pattern of results.

We analysed only large mammals, which by itself should minimize taphonomic or collecting biases. All reconstructed body masses exceed 5 kg and 79 per cent of species are at least 20 kg. There are no temporal trends excluding an increase in body mass of brachydont suoids²³, which, if it causes a bias, might dampen the hypsodonty trends. On the other hand, hypsodont species are 2.8 times the average mass of brachydont species (544 kg versus 197 kg), which, if it causes a bias, might exaggerate the hypsodonty trends. We note that such a bias should not affect the trends calculated separately for brachydont and hypsodont groups. In horses hypsodonty results in a more compact dental design but their exclusion did not eliminate the patterns. Exclusion of Primates and Proboscidea, which could have distinct taphonomic or collecting biases, did not alter the pattern of results.

Although the MN units are distributed in time, the general problem with the application of time-series methods to fossil data is that sampling intervals and efforts vary and samples do not represent single points. Therefore, because we needed a robust measure of the overall trends in the increase of hypsodonty during the Neogene, we examined whether the medians of observed cumulative hypsodonty curves were significantly shifted to younger MN units. Such a shift would indicate a disproportional presence of high hypsodonty values in the younger MN units. We used a randomization test²⁴ to obtain significance values for each median separately by randomly reshuffling the temporal order of MN units 1,000 times and calculating a median for each randomization (medians expected to be clustered around MN 7 + 8 and 9). These null distributions of medians provide a robust measure of trends because only overall increases or decreases depart significantly from the randomizations. For example, common groups had medians

between MN 9.6 and 9.9, indicating an increase in hypsodonty, and we calculated how many times these high, or higher, values were obtained from the random distribution of medians. Because the larger number of relatively brief MN units in the late Miocene may bias the medians, we performed the analyses also after combining MN units 10 and 11, and these P values are given below in parentheses. Our primary interest was not to estimate the absolute increases but to portray trends among common and rare groups, which should be relatively insensitive to differences in MN duration.

The resulting probability values for common (present at ≥ 25 per cent of localities) species, genera and crown types are $P = 0.010, 0.008, 0.008$ (0.017, 0.024, 0.007), and the corresponding values for rare groups are $P = 0.054, 0.371, 1.000$ (0.071, 0.480, 1.000). The crown types have slight decrease in hypsodonty, $P = 0.157$ (0.107), Fig. 2b. The values for common groups excluding hipparionine horses are $P = 0.536, 0.012, 0.048$ (0.598, 0.024, 0.100). The non-significant value for species results from the fact that the cut-off point defining common groups leaves only one common hipparionine species in MN 13. When common and rare groups are combined the increase in hypsodonty is intermediate. The values for species, genera and crown types are $P = 0.025, 0.079, 0.032$ (0.005, 0.299, 0.076).

To test how the relative locality coverage increased in brachydont and hypsodont groups, we examined the shifts in medians of cumulative locality coverage curves. In the few cases where a genus or a crown type had species with a different degree of hypsodonty (see above), we considered these groups hypsodont when their locality averages of hypsodonty were at least 1.5. The probability values obtained from reshuffling MN units for hypsodont species, genera and crown types are $P = 0.017, 0.022, 0.028$ (0.079, 0.065, 0.013), and the corresponding values for brachydont groups are $P = 0.054, 0.148, 0.570$ (0.461, 0.597, 0.553).

Received 26 October 2001; accepted 13 February 2002.

1. Simpson, G. G. *Horses* (Oxford Univ. Press, New York, 1951).
2. Van Valen, L. A functional index of hypsodonty. *Evolution* **14**, 531–532 (1960).
3. Janis, C. M. & Fortelius, M. On the means whereby mammals achieve increased functional durability of their dentitions, with special reference to limiting factors. *Biol. Rev.* **63**, 197–230 (1988).
4. Meng, J. & McKenna, M. C. Faunal turnovers of Palaeogene mammals from the Mongolian plateau. *Nature* **394**, 364–367 (1998).
5. MacFadden, B. J. in *Evolution of Herbivory in Terrestrial Vertebrates* (ed. Sues, H.-D.) 223–244 (Cambridge Univ. Press, Cambridge, 2000).
6. Alroy, J. Constant extinction, constrained diversification, and uncoordinated stasis in North American mammals. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **127**, 285–311 (1996).
7. Alroy, J. et al. Effects of sampling standardization on estimates of Phanerozoic marine diversification. *Proc. Natl Acad. Sci. USA* **98**, 6261–6266 (2001).
8. Damuth, J. Analysis of the preservation of community structure in assemblages of fossil mammals. *Paleobiology* **8**, 434–446 (1982).
9. Fortelius, M., et al. in *The Evolution of Western Eurasian Neogene Mammal Faunas* (eds Bernor, R. L., Fahlbusch, V. & Mittmann, H.-V.) 414–448 (Columbia Univ. Press, New York, 1996).
10. Barnosky, A. D. Distinguishing the effects of the Red Queen and Court Jester on Miocene mammal evolution in the northern Rocky Mountains. *J. Vert. Paleontol.* **21**, 172–185 (2001).
11. Mein, P. in *European Neogene Mammal Chronology* (eds Lindsay, E. H., Fahlbusch, V. & Mein, P.) 73–90 (Plenum, New York, 1989).
12. Steininger, F. E., et al. in *The Evolution of Western Eurasian Neogene Mammal Faunas* (eds Bernor, R. L., Fahlbusch, V. & Mittmann, H.-V.) 7–46 (Columbia Univ. Press, New York, 1996).
13. Alroy, J., Bernor, R. L., Fortelius, M. & Werdelin, L. The MN system: Regional or continental? *Mitt. Bayerischen Staatsammlung Paläontol. Hist. Geol.* **38**, 243–258 (1998).
14. Jernvall, J. Mammalian molar cusp patterns: Developmental mechanisms of diversity. *Acta Zool. Fennica* **198**, 1–61 (1995).
15. Jernvall, J., Hunter, J. P. & Fortelius, M. Molar tooth diversity, disparity, and ecology in Cenozoic Ungulate Radiations. *Science* **274**, 1489–1492 (1996).
16. Garcés, M., Cabrera, L., Agustí, J. & Parés, J. M. Old World first appearance datum of “Hipparion” horses: Late Miocene large-mammal dispersal and global events. *Geology* **25**, 19–22 (1997).
17. Broccoli, A. J. & Manabe, S. in *Tectonic Uplift and Climate Change* (ed. Ruddiman, W. F.) 89–121 (Plenum, New York, 1997).
18. Ruddiman, W. F. (ed.) *Tectonic Uplift and Climate* (Plenum, New York, 1997).
19. An, Z., Kutzbach, J. E., Prell, W. L. & Porter, S. C. Evolution of Asian monsoons and phased uplift of the Himalaya–Tibetan plateau since Late Miocene times. *Nature* **411**, 62–66 (2001).
20. Jarman, P. J. The social organization of antelope in relation to their ecology. *Behavior* **48**, 215–267 (1974).
21. Janis, C. M. Evolution of horns in ungulates: Ecology and palaeoecology. *Biol. Rev.* **57**, 216–317 (1982).
22. Pérez-Barbería, F. J., Gordon, I. J. & Nore, C. Evolutionary transitions among feeding styles and habitats in ungulates. *Evol. Ecol. Res.* **3**, 221–230 (2001).
23. Fortelius, M. & Hokkanen, A. in *Phylogeny of the Neogene Hominoid Primates of Eurasia* (eds De Bonis, L., Koufos, G. & Andrews, A.) (Cambridge Univ. Press, Cambridge, 2001).
24. Manly, B. F. J. *Randomization, Bootstrap and Monte Carlo Methods in Biology* (Chapman and Hall, London, 1997).

Acknowledgements

We thank A. D. Barnosky, J. Damuth, I. Hanski, J. P. Hunter and P. C. Wright for discussions and advice on this work, which was supported by the Academy of Finland.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.J. (e-mail: jivakudaret@aol.com) or M.F. (e-mail: mikael.fortelius@helsinki.fi).

Single-species models for many-species food webs

W. W. Murdoch*, B. E. Kendall†, R. M. Nisbet*, C. J. Briggs‡, E. McCauley§ & R. Bolser*

* Department of Ecology, Evolution and Marine Biology; and † Donald Bren School of Environmental Science and Management, University of California, Santa Barbara, California 93106, USA

‡ Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA

§ Ecology Division, Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4, Canada

Most species live in species-rich food webs; yet, for a century, most mathematical models for population dynamics have included only one or two species^{1–3}. We ask whether such models are relevant to the real world. Two-species population models of an interacting consumer and resource collapse to one-species dynamics when recruitment to the resource population is unrelated to resource abundance, thereby weakening the coupling between consumer and resource^{4–6}. We predict that, in nature, generalist consumers that feed on many species should similarly show one-species dynamics. We test this prediction using cyclic populations, in which it is easier to infer underlying mechanisms⁷, and which are widespread in nature⁸. Here we show that one-species cycles can be distinguished from consumer–resource cycles by their periods. We then analyse a large number of time series from cyclic populations in nature and show that almost all cycling, generalist consumers examined have periods that are consistent with one-species dynamics. Thus generalist consumers indeed behave as if they were one-species populations, and a one-species model is a valid representation for generalist population dynamics in many-species food webs.

Conventional approaches to nonlinear time-series analysis focus on dynamical invariants such as the dimension of the series, which has recently been used to infer the number of strongly interacting

species in a system^{9,10}. These are powerful methods, but require time series much longer than those typically available for field populations¹⁰. In contrast, cycle period can be estimated from relatively short series and, when the organism's maturation time is used to fix the timescale, provides useful dynamical information. Theory for stage-structured populations leads us to consider three classes of cycles. Single-species populations with direct density dependence in vital rates can exhibit 'single-generation cycles' with cycle period one to two times the maturation time^{7,11,12}. Single-species populations can also exhibit 'delayed-feedback cycles' that typically have periods two to four times the maturation time^{12–14} (though longer periods may be possible in models with extremely large-amplitude cycles¹⁵). Models of a specialized consumer, tightly coupled to a resource population so that each controls the dynamics of the other, show longer-period, true 'consumer–resource cycles'. However, approximately constant resource recruitment in these consumer–resource models induces weak coupling and a collapse to single-species dynamics; direct or delayed density dependence in the consumer can then produce single-generation cycles or delayed-feedback cycles in the consumer, with the period determined by consumer development time^{4–6}.

We first establish that specialist consumer–resource cycles can be distinguished from single-species cycles by their scaled periods (cycle period divided by time to maturity). If T_C and T_R are the maturation times of the consumer and resource respectively, then cycles in single-species models have periods that seldom exceed $4T_C$, as noted, whereas consumer resource cycles have periods seldom less than $4T_C + 2T_R$ (Box 1).

Next, the collapse from consumer–resource dynamics to single-species dynamics caused by weak coupling suggests the following prediction. Generalist consumers should typically be weakly coupled to any one of their prey populations because, when feeding on many different species, they cannot be strongly coupled to any one of them. In particular, total resource recruitment rate will be largely independent of the abundance of the consumer and of any particular resource population. We therefore predict that, among cyclic species, generalists should typically show single-species cycles. Specialist species, on the other hand, should more typically show consumer–resource cycles, although strong density dependence

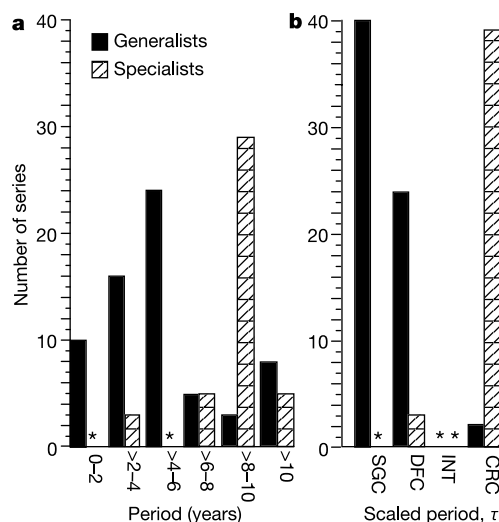


Figure 1 Cycles classified by period. Asterisk indicates zero series in the class. **a**, Number of cyclic populations with various periods, in years. **b**, Distribution of cycles among classes defined by scaled period. SGC, single-generation cycles ($\tau=1$); DFC, delayed-feedback cycles ($2 \leq \tau \leq 4$); CRC, consumer–resources cycles (period in years $\geq 4T_C + 2T_R$). No cycles fall in the intermediate class (INT) between single-species and consumer–resource cycles.

Table 1 Species analysed

Generalists	Specialists
<i>Phylloscopus trochilus</i> , willow warbler	<i>Bupalus piniarius</i> , pine looper (2)
<i>Grus americana</i> , whooping crane	<i>Hyloicus pinastri</i> , pine hawkmoth
<i>Milvus milvus</i> , red kite	<i>Lymantria dispar</i> , gypsy moth
<i>Bucephala albeola</i> , bufflehead	<i>Zeiraphera diniana</i> , larch budmoth
<i>Parus major</i> , great tit	<i>Dendrolimus pini</i> , pine-tree lappet
<i>Esox lucius</i> , pike	<i>Lymantria monacha</i> , black arches
<i>Rutilus rutilus</i> , roach (2)	<i>Epirrita autumnata</i> , autumnal moth
<i>Coregonus albula</i> , vendace	<i>Exapate duratella</i>
<i>Perca flavescens</i> , yellow perch	<i>Panolis flammea</i> , pine beauty
<i>Merlangius merlangius</i> , whiting	<i>Lepus americanus</i> , snowshoe hare
<i>Oncorhynchus gorbis</i> , pink salmon (8)	<i>Lepus europaeus</i> , brown hare
<i>Oncorhynchus nerka</i> , sockeye salmon (21)	<i>Lepus timidus</i> , mountain hare (2)
<i>Salmo salar</i> , Atlantic salmon	<i>Lynx canadensis</i> , Canadian lynx (15)
<i>Pleuronectes platessa</i> , plaice	<i>Mustela vison</i> , North American mink (10)
<i>Gadus morhua</i> , cod (3)	<i>Onodra zibethicus</i> , muskrat (3)
<i>Clupea harengus</i> , Atlantic herring	
<i>Castor canadensis</i> , beaver (4)	
<i>Taxidea taxus</i> , American badger	
<i>Ursus americanus</i> , black bear (3)	
<i>Sus scrofa</i> , wild boar (2)	
<i>Ovis montanus</i> , muskox (2)	
<i>Cancer magister</i> , Dungeness crab (5)	
<i>Calathus melanocephalus</i> , carabid beetle	
<i>Pterostichus versicolor</i> , carabid beetle	
<i>Vespa sp.</i> , wasp	

Except as noted, generalists had single-species cycles and specialists had consumer–resource cycles. Species were chosen based on match between life-history features and discrete-time theoretical models (Methods and Supplementary Information 3). All appropriate series for a species were analysed (Methods) and, if this exceeded 1, the number of series analysed is in parenthesis. *L. monacha* and two of the mink series showed single-species cycles. Each of the two carabid beetle species showed consumer–resource cycles.

Box 1

A lower bound for periods of consumer–resource cycles

A family of consumer–resource models has the property that periodic or quasi-periodic solutions have period greater than six¹⁶. These models assume that both the consumer and resource species reproduce only once, and have the same development time. Here we generalize the work of ref. 16 to include species with arbitrary development times and show that, with biologically reasonable parameter values, appropriately scaled periods are seldom likely to drop much below six, even when the strict mathematical conditions for that bound are relaxed.

We consider a tightly coupled consumer–resource model of the general form:

$$R_t = \lambda R_{t-1} f(R_{t-1}, C_{t-1}) + S_R R_{t-1} \quad (1)$$

$$C_t = \lambda R_{t-1} (1 - f(R_{t-1}, C_{t-1})) + S_C C_{t-1} \quad (2)$$

where R_t and C_t represent the density of adults in the resource and consumer populations in year t , λ is the geometric rate of resource increase, $f(R_t, C_t)$ is the fraction of the resource population that survives attack by the consumer, and S_R , S_C are the fractions of the resource and consumer populations surviving to the next year.

We assume that the system has only one non-trivial equilibrium state (R^* , C^*). The characteristic equation defining the conditions for local stability of the equilibrium is:

$$\mu^2 + A_1 \mu + A_2 = 0 \quad (3)$$

with $A_1 = -1 - \alpha - \beta - S_C$, $A_2 = \alpha(\lambda + S_R) + S_C(1 + \beta)$, where $\alpha = -\lambda R^* f_C(R^*, C^*)$, $\beta = \lambda R^* f_C(R^*, C^*)$, f_i implies a partial derivative with respect to i for $i = R$ or C and μ is an eigenvalue. With forms of the functional response most appropriate for specialist consumers (types 1 or 2), α and β are both non-negative.

On the local stability boundary, $A_2 = 1$, and the period of the cycles is

$$P = 2\pi / \tan^{-1} \left(\sqrt{\frac{4}{(1 + \alpha + \beta + S_C)^2} - 1} \right) \quad (4)$$

which always exceeds six. With some additional algebra, it can be shown that the cycle period approaches six for large λ only if $\beta = 0$ and $S_C = 0$. If either $\beta > 0$ (as with a type 2 functional response) or $S_C > 0$ (adult consumers may live for more than a year), the limiting period is greater than six. Cycle period increases as λ decreases from the limit to lower, and biologically realistic, values.

Resource density dependence can reduce the period, but only slightly. We analysed a variant of the above model with the standard Nicholson–Bailey assumption, $f = \exp(-\alpha C_t)$, implying $\beta = 0$, but with either Ricker or Beverton–Holt density dependence in the resource population. A period of six is no longer a lower bound, but we found that consumer–resource periods always exceeded 5.5 with $\lambda < 10$, typically the upper limit in natural populations (Supplementary Information 3). (Delayed-feedback cycles (a period of two) arise in the Ricker version with sufficiently strong resource density dependence.)

A generalization of model (1) that allows for different juvenile development times, T_R and T_C , for resource and consumer respectively, has equations:

$$R_t = \lambda R_{t-T_R} f(R_{t-T_R}, C_{t-T_R}) + S_R R_{t-1} \quad (5)$$

$$C_t = \lambda R_{t-T_C} (1 - f(R_{t-T_C}, C_{t-T_C})) + S_C C_{t-1} \quad (6)$$

With the same restrictions as before ($\beta = 0$ and $S_C = 0$), and one additional assumption ($S_R = 0$), we can prove that the cycle period approaches a limiting value of $2T_R + 4T_C$, as $\lambda \rightarrow \infty$. An outline of the proof is in Supplementary Information 1. We have not managed to prove that this limit is a lower bound as was true with a period of six, but extensive numerical studies suggest that it is.

The above analyses are strictly valid only on the stability boundary. However, such calculations are typically a good indicator of the periods of all but very large-amplitude quasi-cycles or limit cycles. With large-amplitude limit cycles, periods are typically longer than would be predicted from our analysis¹⁷, supporting our interpretation of $2T_R + 4T_C$ as a lower bound.

might sometimes reduce these to single-species cycles.

We analysed 108 time series from natural populations (comprising 40 species), each of which exhibited statistically significant cycles and met criteria ensuring a rigorous test of the predictions (Methods). We classified each species into one of two trophic roles (Methods). ‘Generalist’ consumers (for example, predatory fish) feed on many resource species over their lifetime and in turn are not preyed on by specialist consumers. ‘Specialists’ (for example, lynx) feed largely on one species, or are prey of a specialist predator (for example, snowshoe hares). We then estimated cycle periods and calculated scaled periods. Our focus was on generalists, because the specialists in the database are known, in general, to have long-period cycles. We used the distribution of specialist cycle periods as a basis for comparison and to test the Box 1 prediction that scaled long periods are not less than $4T_C + 2T_R$.

Our predictions are well supported. The generalists and specialists did not separate cleanly when we examined period measured in years (mainly because a substantial fraction of generalist periods are longer than four, or even six, years; see Fig. 1a) but they did so almost perfectly when we used the scaled period (Fig. 1b). All but two of the 66 generalist consumer series show single-species cycles, including those with periods greater than four years. There are no cycles with intermediate scaled period, that is, between $4T_C$ and $4T_C + 2T_R$, in spite of the fact that many periods lie between four and six years (Fig. 1a). As expected, the great majority of cyclic series from specialists are true consumer–resource cycles (Fig. 1a), and none of these periods is less than $4T_C + 2T_R$, as predicted by Box 1. Furthermore, as predicted by the theory discussed above, when specialists have shorter periods (only three cases), they are single-species periods rather than intermediate periods (Fig. 1b).

Single-species cycles occur in a broad taxonomic range but are especially prevalent in fish (Table 1). True consumer–resource cycles are restricted in this data set to forest Lepidoptera and mammalian carnivores and their prey, but probably occur in other taxonomic groups that were excluded because the species have multiple generations per year.

Our results suggest not only that single-species models are often appropriate for describing population dynamics of individual species in many-species food webs, but that prey species diversity, a prevalent feature of natural systems, itself facilitates such simplification. Remaining challenges include adding taxonomic variety to the database, developing appropriate theory on cycle periods for multivoltine species, and determining whether the simplification applies in non-cyclic generalist populations. □

Methods

The time series are from the Global Population Dynamics Database¹⁸, plus the three additional series found while searching the literature for relevant natural history information (Supplementary Information 2). Criteria for excluding time series of field populations from the analysis (see Supplementary Information 3) and the main groups excluded, are as follows:

(1) We could not determine the species’ trophic role (nine series from insects and game birds).

(2) There is strong evidence the population has no causal role in the cycle (43 series from high-latitude species, mainly foxes, entrained by the lynx–snowshoe hare cycle or, less often, by microtine (voles, lemmings, and so on) cycles; these populations track but do not cause the cycles^{19–21}). Mink and muskrat series were analysed because there is evidence that they are causally involved in a distinct cycle¹⁹, as were the mountain and brown hare series²².

(3) To conform to the discrete-time theory, series were excluded if maturation takes less than one year and breeding is more frequent than annual. Theory on cycle periods is not available for seasonally driven, properly age-structured, consumer–resource models of species with more than one generation per year. This excludes 31 series, mainly from voles.

(4) For statistical power, we focused on time series at least 25 years long, and excluded any results in which the estimated period was more than one-third of the series length. To avoid biologically meaningless results, we only analysed data with mean counts of more than 10 individuals and cycles in which the maximum population was on average at least twice the minimum.

Information from the literature on trophic role was supplemented by discussions with appropriate experts. Generalist consumers—usually top predators with broad diets—could typically be categorized easily and unambiguously. Lepidopteran forest pests were classified as specialist prey (of parasitoids)²³, although some cycles may be specialized

pest–plant cycles or even single-species dynamics with second-order delays such as maternal effects, which lead to periods of over $6T$ (ref. 24). The last would increase the prevalence of single-species dynamics.

Scaled period, τ , was calculated by dividing the period (in years) by time to maturity (in years), T_C (sources in Supplementary Information 3). For cases in which $\tau > 4$, we then asked if the cycle period in years exceeded $4T_C + 2T_R$; in series involving lynx (average maturation time = 1.5 years) this sum was eight years; in all others, where consumer and resource species develop in one year, it was six. Because maturation times and cycle periods are approximations, estimates of τ were rounded to the nearest integer. Periods with $\tau = 1$ were assigned to single-generation cycles; those with $2 \leq \tau \leq 4$ were assigned to delayed-feedback cycles.

The series were analysed 'blind', without knowledge of the species name or trophic role. We used multitaper spectral analysis^{25,26} using the algorithm described in ref. 27. This technique avoids the often arbitrary choice of data tapering by using a series of optimal Slepian tapers. We classified a series as periodic if there was both a clear peak in the power spectrum (of either the transformed or the untransformed data) and the corresponding P -value was <0.05 . If there were missing values, we analysed the longest continuous segment of the data. If, upon visual inspection, only part of the series appeared cyclic, we analysed that segment. We analysed the data both with and without a logarithmic transformation, and removed any linear trend in both cases. If several series were available from any given area and time period, we analysed only one series, and only the predator if predator and prey series were available.

Received 30 October 2001; accepted 5 March 2002.

1. Lotka, A. J. *Elements of Physical Biology* (Dover, New York, 1925, reprinted 1956).
2. Hassell, M. P., Lawton, J. H. & May, R. M. Patterns of dynamical behaviour in single species populations. *J. Anim. Ecol.* **45**, 471–486 (1976).
3. Gurney, W. S. C. & Nisbet, R. M. *Ecological Dynamics* (Oxford Univ. Press, New York, 1998).
4. McCauley, E., Nisbet, R. M., De Roos, A. M., Murdoch, W. W. & Gurney, W. S. C. Structured population models of herbivorous zooplankton. *Ecol. Monogr.* **66**, 479–501 (1996).
5. Persson, L. *et al.* Ontogenetic scaling of foraging rates and the dynamics of a size-structured consumer-resource model. *Theor. Popul. Biol.* **54**, 270–293 (1998).
6. Briggs, C. J., Nisbet, R. M. & Murdoch, W. W. Host age-specific parasitoid gain, delayed-feedback, and multiple attractors in a host-parasitoid model. *J. Math. Biol.* **38**, 317–345 (1999).
7. Kendall, B. E. *et al.* Why do populations cycle? A synthesis of statistical and mechanistic modeling approaches. *Ecology* **80**, 1789–1805 (1999).
8. Kendall, B. E., Prendergast, J. & Bjornstad, O. N. The macroecology of population dynamics: taxonomic and biogeographic patterns in populations cycles. *Ecol. Lett.* **1**, 160–164 (1998).
9. Stenseth, N. C., Falck, W., Bjornstad, O. N. & Krebs, C. J. Population regulation in snowshoe hare and Canadian lynx: Asymmetric food web configurations between hare and lynx. *Proc. Natl Acad. Sci. USA* **94**, 5147–5152 (1997).
10. Bjornstad, O. N., Sait, S. M., Stenseth, N. C., Thompson, D. J. & Begon, M. The impact of specialized enemies on the dimensionality of host dynamics. *Nature* **409**, 1001–1006 (2001).
11. Gurney, W. S. C. & Nisbet, R. M. Fluctuation periodicity, generation separation, and the expression of larval competition. *Theor. Popul. Biol.* **28**, 150–180 (1985).
12. Jones, A. E., Nisbet, R. M., Gurney, W. S. C. & Blythe, S. P. Period to delay ratio near stability boundaries for systems with delayed feedback. *Math. Anal. Appl.* **135**, 354–368 (1988).
13. Blythe, S. P., Nisbet, R. M. & Gurney, W. S. C. Instability and complex dynamical behaviour in population models with long time delays. *Theor. Popul. Biol.* **22**, 147–176 (1982).
14. Higgins, K., Hastings, A. & Botsford, L. W. Density dependence and age structure: Nonlinear dynamics and population behaviour. *Am. Nat.* **149**, 247–269 (1997).
15. Nisbet, R. M. & Bence, J. R. Alternative dynamic regimes for canopy-forming kelp: A variant on density-vague population regulation. *Am. Nat.* **134**, 377–408 (1989).
16. Lauwerier, H. A. & Metz, J. A. J. Hopf bifurcation in host-parasitoid models. *Inst. Math. Appl. J. Math. Appl. Med. Biol.* **3**, 191–210 (1986).
17. Nisbet, R. M. & Gurney, W. S. C. *Modelling Fluctuating Populations* (Wiley, New York, 1982).
18. NERC Centre for Population Biology, Imperial College The Global Population Dynamics Database, (NERC, Ascot 1999); available at <http://www.sw.ic.ac.uk/cpb/gpdd.html>.
19. Finerty, J. P. *The Population Ecology of Cycles in Small Mammals: Mathematical Theory and Biological Fact* (Yale Univ. Press, New Haven, 1980).
20. Lindstrom, E. R. *et al.* Disease reveals the predator: Sarcophagid mange, red fox predation, and prey populations. *Ecology* **75**, 1042–1049 (1994).
21. O'Donoghue, M., Boutin, S., Hofer, E. J. & Boonstra, R. In *Ecosystem Dynamics of the Boreal Forest: the Kluane Project* (eds Krebs, C. J., Boutin, S. & Boonstra, R.) 325–340 (Oxford Univ. Press, New York, 2001).
22. Bulmer, M. G. A statistical analysis of the 10-year cycle in Canada. *J. Anim. Ecol.* **43**, 701–718 (1974).
23. Berryman, A. A. What causes population cycles of forest Lepidoptera? *Trends Ecol. Evol.* **11**, 28–32 (1996).
24. Ginzburg, L. R. & Taneyhill, D. E. Population cycles of forest Lepidoptera: a maternal effect hypothesis. *J. Anim. Ecol.* **63**, 79–92 (1994).
25. Thomson, D. J. Spectrum estimation and harmonic analysis. *Proc. IEEE* **70**, 1055–1096 (1982).
26. Park, J., Lindberg, C. R. & Vernon, F. L. Multitaper spectral analysis of high-frequency spectrograms. *J. Geophys. Res.* **92**, 12675–12684 (1987).
27. Lees, J. M. & Park, J. Multiple-taper spectral analysis: a stand-alone C-subroutine. *Comput. Geosci.* **21**, 199–236 (1995).

Acknowledgements

We thank C. Godfray for permission to analyse time series from the Global Population Dynamics Database (GPDD). We thank J. Fryxell, B.K. Gilbert, S. Henke, P.J. Hudson, C.J. Krebs, L. Oksanen, B. Sanderson, S. Taylor and M. Tewes for discussions on particular taxa, and J.A.J. Metz for discussions of cycle periods. The research was supported by grants from the NSF and United States Department of Agriculture (USDA).

Competing interests statement

The authors declare that they have no competing financial interests

Correspondence and requests for materials should be addressed to W.M. (e-mail: murdoch@lifesci.ucsb.edu).

Precise inhibition is essential for microsecond interaural time difference coding

Antje Brand*, Oliver Behrend*, Torsten Marquardt†, David McAlpine† & Benedikt Grothe*

* Max Planck Institute of Neurobiology, Am Klopferspitz 18a, 82152 Martinsried, Germany

† Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK

Microsecond differences in the arrival time of a sound at the two ears (interaural time differences, ITDs) are the main cue for localizing low-frequency sounds in space. Traditionally, ITDs are thought to be encoded by an array of coincidence-detector neurons, receiving excitatory inputs from the two ears via axons of variable length ('delay lines'), to create a topographic map of azimuthal auditory space^{1,2}. Compelling evidence for the existence of such a map in the mammalian ITD detector, the medial superior olive (MSO), however, is lacking. Equally puzzling is the role of a—temporally very precise³—glycine-mediated inhibitory input to MSO neurons. Using *in vivo* recordings from the MSO of the Mongolian gerbil, we found the responses of ITD-sensitive neurons to be inconsistent with the idea of a topographic map of auditory space. Moreover, local application of glycine and its antagonist strychnine by iontophoresis (through glass pipette electrodes, by means of an electric current) revealed that precisely timed glycine-controlled inhibition is a critical part of the mechanism by which the physiologically relevant range of ITDs is encoded in the MSO. A computer model, simulating the response of a coincidence-detector neuron with bilateral excitatory inputs and a temporally precise contralateral inhibitory input, supports this conclusion.

Sensitivity to ITDs, which can be in the order of only micro-seconds⁴, requires neuronal processing with a temporal acuity far beyond that normally observed in the mammalian central nervous system. Traditionally, ITD processing in mammals and birds is explained by means of an elegant model devised by Jeffress more than 50 years ago¹. The model assumes the existence of arrays of coincidence-detector neurons receiving excitatory inputs from the two ears. Neurons respond maximally when stimulus-evoked action potentials, phase-locked to the stimulus waveform, converge from each ear simultaneously. Different conduction delays from each ear, assumed to result from a system of delay lines, provide a means by which different coincidence-detector neurons encode different ITDs. A systematic arrangement of such delay lines is assumed to create a topographic representation of ITDs and, thus, a map of sound positions in the azimuthal plane.

For nucleus laminaris, the ITD-processing structure in birds, the existence of such an arrangement has been confirmed^{5,6}. However, for the mammalian equivalent, the MSO (Fig. 1a), direct confirmation of such an arrangement is lacking. Spherical bushy cells from each cochlear nucleus provide binaural excitatory input to MSO neurons^{7,8}, which *in vivo* recordings demonstrate to be sensitive to ITDs^{9,10}. However, MSO neurons also receive prominent

glycine-containing inputs directly onto their somata, arising mainly from the medial, and partially from the lateral, nucleus of the trapezoid body (MNTB, LNTB; Fig. 1a)^{11–14}. The contribution of these inputs to ITD processing remains unknown^{3,15}. To test directly the extent to which the established Jeffress model is relevant to the processing of ITDs in mammals, we recorded responses of ITD-sensitive neurons from the MSO of the Mongolian gerbil (*Meriones unguiculatus*), a rodent with well-developed low-frequency hearing. Mongolian gerbils use ITDs for low-frequency sound localization¹⁶, and have MSO neurons that show similar ITD sensitivity to that reported for cats and dogs¹⁰. Ionophoretic application of glycine and its antagonist strychnine permitted the assessment of the role of inhibition in the mammalian ITD detector.

Responses were recorded from 36 low-frequency MSO neurons, of which 24 were binaurally excited. Of these, 20 neurons were sensitive to ITDs with 30–85% change in the discharge rate across the physiologically relevant range of ITDs of the gerbil (roughly $\pm 120 \mu\text{s}$; shaded area in Fig. 1b). Most neurons (16 of 20) responded selectively to ITDs of pure tones, 4 to ITDs of only broad-band stimuli. The response maximum (best ITD) of individual neurons was largely independent of the stimulus frequency tested (Fig. 1b), a feature consistent with the Jeffress model⁹. All neurons responded maximally to sounds leading in time at the contralateral ear (positive ITDs). Note that for the neuron in Fig. 1b, maximal response is evoked by an ITD of $+200 \mu\text{s}$, outside the physiologically relevant range of the gerbil. This was a general observation.

According to a commonly held view of the Jeffress model, ITDs are systematically represented by the activity of individual coincidence-

detector neurons, creating a place code of azimuthal position by virtue of their maximum firing rates^{1,2}. It is generally assumed in this scheme that response maxima of ITD functions are independent of the frequency-tuning characteristics of the neurons. However, our data do not support this view. First, plotting the response of each MSO neuron to interaurally delayed 'best-frequency' tones indicate that most neurons respond maximally outside the physiologically relevant range ($n = 16$; Fig. 2a). As a consequence, ITD functions are steepest within the physiologically relevant range, creating a maximum (and unambiguous) dynamic range for each neuron's ITD-evoked response within this range. Furthermore, the ITD that evokes the peak response is dependent on neuronal tuning for sound frequency (Fig. 2b); neurons with relatively low best frequencies respond maximally at relatively long ITDs, whereas neurons with relatively high best frequencies respond maximally at relatively short ITDs, although with slight scatter of ITD maxima of neurons with best frequencies higher than 700 Hz. This correlation between best frequency and peak ITD, which is consistent with a report from a much larger sample of neurons in the auditory midbrain¹⁷, is highly significant ($r^2 = 0.83$). When expressed in terms of the equivalent best interaural phase difference (IPD), that is, normalized for each neuron's best frequency, the peak response occurs at a constant IPD of $+0.12$ cycles (s.d. ± 0.043 ; $n = 16$). Consequently, the map of best ITDs in the MSO appears to run along the tonotopic gradient; that is, there appears to be no 'place map' of ITDs within individual frequency channels.

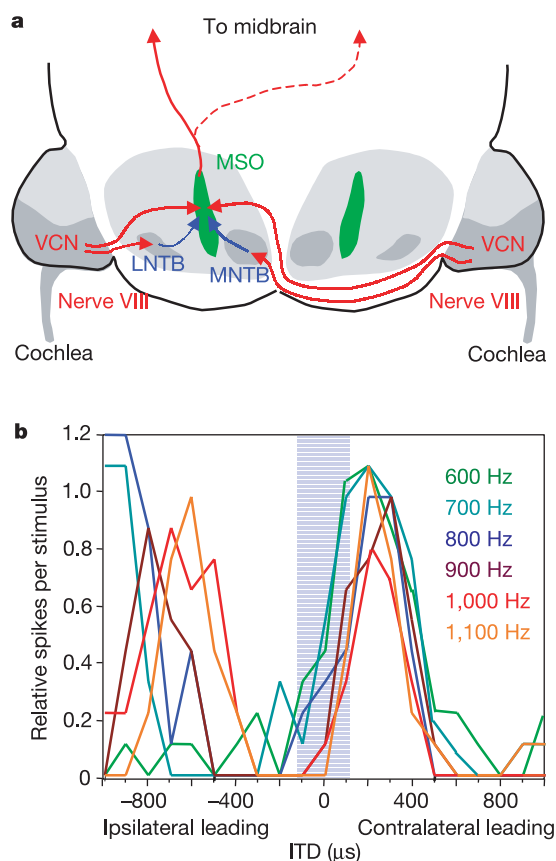


Figure 1 ITD coding in the mammalian MSO. **a**, Innervation pattern of the MSO. Excitatory inputs (red) from ventral cochlear nucleus (VCN), and glycinergic inputs (blue) from MNTB and LNTB are shown. **b**, ITD function of a gerbil MSO neuron, tested with pure tones at different frequencies. As is typical, the peak ITD is independent of stimulus frequency. The blue area indicates the physiologically relevant range for gerbils ($\pm 120 \mu\text{s}$).

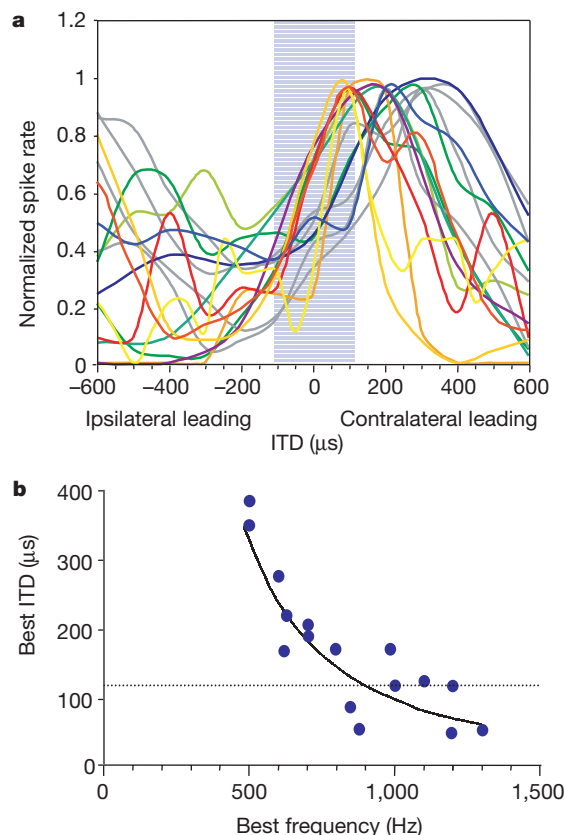


Figure 2 Peaks of ITD functions are outside of the physiologically relevant range. **a**, ITD functions of gerbil MSO neurons, each tested at its best frequency, 20 dB above threshold. Note that the peaks are largely outside the physiologically relevant range (blue area). The steep slopes of the ITD functions are within the relevant area. **b**, Distribution of best ITDs as a function of best frequencies of the neurons. Only ITDs below the dotted line are within the physiologically relevant range of ITDs. Solid line, third-order polynomial fit. Dotted line, upper limit of the gerbil's physiological range of ITDs.

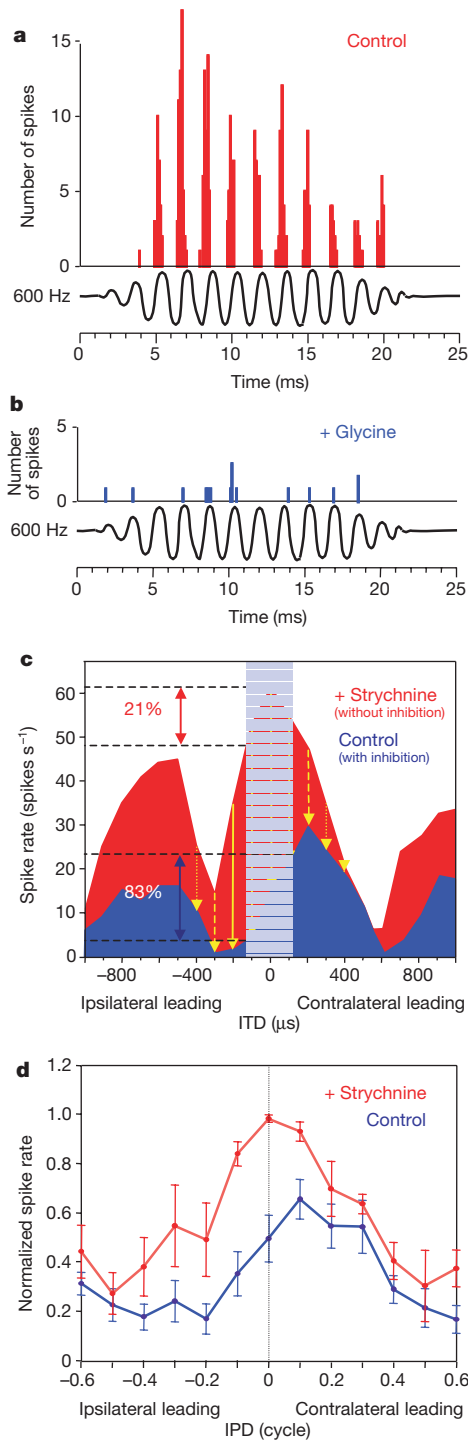


Figure 3 Effects of glycine and its antagonist strychnine in gerbil MSO neurons **a**, Peri-stimulus-time histogram showing precise phase-locking of a neuron in response to a 600-Hz pure tone. **b**, Iontophoretic application of glycine with minimal currents (<30 nA) and for only a few seconds completely inhibits the tone-evoked response. **c**, ITD functions of a neuron (best frequency 1,060 Hz) during control condition and strychnine application. Dynamic range of the ITD response under control conditions (blue) is 83%; the best ITD is outside of the physiologically relevant range (blue area). During blockade of glycinergic inhibition from application of strychnine (red), the best ITD is at zero ITD, and the dynamic range of the response is reduced to 21%. Hence, glycinergic inhibition reduces spike counts and shifts the left slope into the relevant range of ITDs. The yellow arrows indicate the different effects of inhibition at different ITDs. **d**, Averaged interaural delay functions (ITDs converted to IPD, thereby normalizing for BF) for all five ITD-sensitive neurons tested under control conditions (blue) and during application of strychnine (red). Before averaging, each neuron's IPD function was normalized to the maximal response during strychnine application (see text). Error bars, s.e.m.

The most puzzling finding in the mammalian ITD encoding circuitry is the glycinergic inhibitory input to the MSO. To determine the role of this input on ITD coding, the inhibitory transmitter glycine ($n = 5$) or its antagonist strychnine ($n = 6$) was applied during recordings made from nine MSO neurons (two neurons were successfully tested for both). For all five neurons, application of glycine almost completely abolished the response to the test tone (Fig. 3a, b) confirming the existence of glycine receptors and their profound inhibitory effect on MSO neurons.

To assess more directly the influence of glycinergic inhibition on ITD sensitivity of individual MSO neurons, we successfully applied the glycine antagonist strychnine on six neurons. Five of these neurons were sensitive to ITDs. Two effects were observed. First, in all six neurons, strychnine produced a significant increase in discharge rate (50–86% at or close to zero ITD). Second, for each of the five neurons that were ITD sensitive, application of strychnine shifted the peak of the ITD functions towards zero ITD. The blue area in Fig. 3c shows the ITD function for one of the five neurons (best frequency 1,060 Hz) under control conditions (in the absence of strychnine). The ITD function showed the typical monotonic increase of discharge rate across the relevant extent of ITDs, in this case of about 83%. The neuron's peak ITD was $+170$ μ s, outside the physiologically relevant range for the gerbil. When strychnine was applied (red area in Fig. 3c), the best ITD shifted downwards to $+50$ μ s, some 120 μ s from its initial peak, corresponding to a shift in best IPD of 0.127 cycles. The steepest slope of the function therefore fell outside the physiologically relevant range of ITDs. In addition, the function became nonmonotonic within the physiological range and, hence, ambiguous for ITD. The dynamic range within the physiological range was reduced from 83% to 21%. In each of the five neurons tested, strychnine shifted the best ITD close to zero ITD. The mean equivalent IPD of these five neurons before application of strychnine was 0.148 cycles (± 0.054) and 0.021 cycles (± 0.053) during application of strychnine. Hence, strychnine caused an average shift of -0.127 cycles (± 0.064). The shifts observed in all five neurons were statistically significant (paired t -test, $P < 0.014$).

The averaged IPD functions for the five neurons after first normalizing each function to the maximum discharge rate in the presence of strychnine are shown in Fig. 3d. The average function is similar to that of the example neuron in Fig. 3c. On average, iontophoresis of strychnine produced a 67% ($\pm 9.8\%$) reduction of the dynamic range of the response across the physiologically relevant range of ITDs. The yellow arrows in Fig. 3c indicate the apparent effect of the glycinergic inhibition, which differed at different ITDs, being stronger when the ipsilateral stimulus was leading in time than when the contralateral stimulus was leading. Because glycinergic inputs to the MSO are insensitive to ITDs, the only explanation for this observation is that the inhibition occurs in specific phase relation to the excitatory inputs; that is, it is precisely phase-locked. To explain the influence of inhibition in shifting the peaks of ITD functions in the observed direction, one could assume ipsilaterally driven inhibition that follows (phase lags) the ipsilateral excitation, contralateral inhibition that precedes contralateral excitation, or a combination of both. However, contralateral inhibition preceding contralateral excitation is a common phenomenon in the bat MSO¹⁸, and the contralateral inhibition has been shown to be more prominent than the ipsilateral inhibition^{13,15}.

The shift in ITD tuning that results from a phase-locked, contralateral inhibition preceding contralateral excitation was simulated using a modified Hodgkin–Huxley model¹⁹ (Fig. 4) on the basis of physiological findings in neurons of the anteroventral cochlear nucleus (AVCN). AVCN neurons exhibit similar nonlinear membrane properties to those seen in the MSO. In this model, an additional voltage-dependent, low-threshold potassium channel renders the single-compartment model particularly sensitive to

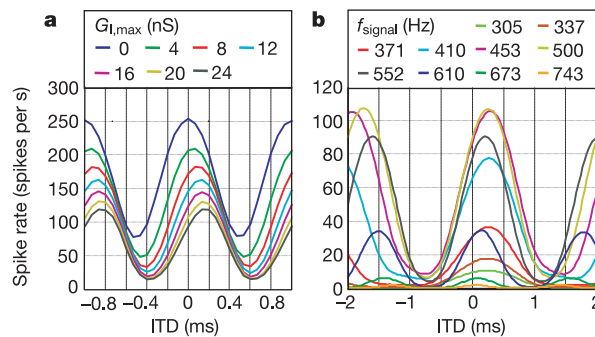


Figure 4 Simulation of the shift in ITD caused by timed contralateral inhibition using a modified Hodgkin–Huxley model (see text). **a**, ITD functions with various inhibitory synaptic strength (tuning maximum synaptic conductance, $G_{l,max}$) show the influence of the contralateral inhibition on best ITD and overall spike rate (signal frequency is

1,000 Hz). **b**, ITD functions of a simulated 500-Hz neuron to interaurally delayed tones of various frequencies. The central peak of the ITD functions is largely independent of the stimulus frequency ($G_{l,max} = 10$ nS).

the relative timing of its spike inputs. Consistent with our physiological observations (Fig 3c, d), Fig. 4a illustrates how increasing the strength (that is, conductance $G_{l,max}$) of the timed inhibition progressively shifts the central peak of the function towards longer values of ITD, and reduces the overall discharge rate. To simulate the physiological results from the stimulation frequency of 1,000 Hz (Fig. 3c), we set both the excitatory and the inhibitory synaptic time constants to $\tau = 0.1$ ms, giving a half-time of 0.25 ms (for lower test frequencies, slower time constants were sufficient). ITD functions for several stimulus frequencies are shown in Fig. 4b, for a simulated neuron with a best frequency of 500 Hz. The typical pattern of a coincidence-detector neuron is evident, with the position of the central peak largely independent of stimulus frequency, although there is a systematic tendency for the peak to shift towards lower ITDs as input level decreases, owing to peripheral auditory filtering. These results indicate that a phase-locked and precisely timed contralateral inhibition, such as innervates the MSO from the MNTB, and preceding the contralateral excitation, can account for the neural responses observed.

Our data indicate that ITD tuning in the gerbil MSO is the product of an interaction of precisely timed excitatory and inhibitory inputs. Moreover, the glycinergic inhibition appears responsible for tuning the steep slope of ITD functions to the physiologically relevant range by suppressing the response to ITDs at which the ipsilateral stimulus is leading. Because the inhibitory input is not ITD sensitive itself, the most likely explanation is that a monaural phase-locked inhibitory input is responsible for this behaviour. In fact, sources of inhibition to the MSO, from the MNTB in particular, show unique morphological and physiological specializations (including calyx synapses) that enable highly temporally precise transmission of neural signals^{20,21}, and strong evidence exists that the contralateral, MNTB-derived inhibition is the dominant inhibitory input to the MSO^{13,15}. To influence responses to negative ITDs (ipsilateral leading), such contralateral inhibition must precede contralateral excitation. Glycinergic inhibition preceding excitation evoked from the same ear is a common phenomenon in the bat MSO^{18,22}. Exact temporal adjustment of the input to the MSO from the MNTB is, therefore, crucial for creating the ‘correct’ ITD, as is an inhibitory time constant much faster than that observed in the rat MSO¹⁴. The temporal adjustment of the fast time constant, however, might be related to the experience-dependent elimination of glycinergic synapses on MSO neurons during the first days after hearing onset, refining the inputs to the MSO somata only—in contrast to the diffuse distribution in the rat²³. The finding that inhibitory transmitters are crucial in processing ITDs (as with other temporal cues⁸), taken together with evidence for an age-related downregulation of inhibitory transmitters²⁴, offers new ways of explaining age-related hearing deficits, such as the deterioration of the capacity to

segregate sound sources. Finally, our results provide support for a model of spatial hearing, based on recent observations in the guinea pig midbrain¹⁷, in which the slopes of ITD functions, rather than their peaks, are the critical feature in determining how they might contribute to the localization of sound sources in azimuth. Altogether, the traditional model of low-frequency spatial hearing¹ seems not to be an adequate description of how ITDs are either encoded or represented in the mammalian auditory system. □

Methods

Sample sizes

We recorded from 36 low-frequency (<2 kHz) neurons histologically localized to the MSO, of which 20 showed a profound ITD sensitivity. It is commonly reported to be difficult to record from MSO neurons^{25,9,10}. Additionally, application of glycine and particularly of its antagonist strychnine turned out to be difficult, presumably because of the high density of cell bodies in the central region of the MSO and of the densely packed myelinated MSO input and output fibres that prevented the relatively large molecule from reaching the receptors of the cells from which we recorded. Thus, only ‘ideal’ recording configuration allowed successful application of drugs, which explains the $n = 9$ neurons with drug application in this study.

Delivery of stimuli to animals

Animals (Mongolian gerbils, *Meriones unguiculatus*) were anaesthetized with a mixture of Ketamine (10 mg per 100 g) and Rompun (2%) throughout the entire experiment. During recordings, the animal was placed on a heating cushion in a sound-attenuated chamber. Recording electrodes were inserted stereotaxically through the foramen magnum¹⁰. We used glass pipettes for recordings filled with 2 M sodium acetate (impedance 10–30 M Ω) glued to a five-barrel glass electrode²⁶, allowing local iontophoretic application of 0.5 M glycine (10–40 nA), 0.01 M strychnine (50–250 nA), or 4% solution of horseradish peroxidase (HRP, Sigma) at the recording site. One barrel filled with sodium acetate (1 M) served as a balancing channel. Recording sites were anatomically confirmed by histological analysis of the HRP injections.

Action potentials were recorded using conventional methods described elsewhere²⁷. Only action potentials from single neurons with a signal-to-noise ratio greater than five were recorded.

Stimuli were delivered using Tucker Davis Technologies (TDT) System II²⁷ and two Beyer dynamics speakers (model DT 990), fitted to the ears by probe tubes (2 mm inner diameter). The stimulus-delivering system and the speakers were calibrated using a 1/4-inch microphone (Reinstorp VtS), a measuring amplifier (MV 302, Microtech) and a waveform analyser (Stanford Research Systems, model SR770 FFT network analyser).

Data recording

After electrophysiological isolation of a single neuron, a frequency-tuning curve was measured and the best frequency and threshold determined. Sensitivity to static ITDs was then tested using different stimulus frequencies. For all tests during drug applications, the frequency was set to the best frequency of the neuron, 20 dB above its threshold. Stimulus duration was 100 or 200 ms, delivered at a rate of 4 stimuli s per s with at least ten repetitions. The stimulus delay was constant at the contralateral ear and varied at the ipsilateral, covering a range of up to ± 1 ms (resolution of 50 or 100 μ s). Data were analysed offline using individual spike times and peri-stimulus-time histograms. Also, we performed a vector analysis to evaluate the degree of phase-locking and the best ITD^{25,28}. All experiments were approved according to the German Tierschutzgesetz.

Model parameters

The model is identical to one previously applied in the simulation of MSO neurons²⁹, and, if not otherwise stated, all parameters used here were identical, except that both the excitatory and the inhibitory synaptic time constants were short ($\tau = 0.1$ ms), giving a

half-time of 0.25 ms. The maximum excitatory conductance was $G_{E,max} = 1.5$ nS. All 2×24 excitatory and 24 inhibitory pre-synaptic inputs were generated independently by an auditory nerve model³⁰ that produced about 160 spikes in response to the 1-s-tone-burst stimulus when the tone's frequency equals the best frequency. The contralateral inhibition led the excitation from either side by 0.2 ms. For simplicity, the simulations of cochlear nucleus neurons and inhibitory interneurons in the trapezoid body were omitted.

Received 31 December 2001; accepted 13 March 2002.

- Jeffress, L. A. A place theory of sound localization. *J. Comp. Physiol. Psychol.* **41**, 35–39 (1948).
- Joris, P. X., Smith, P. H. & Yin, T. C. Coincidence detection in the auditory system: 50 years after Jeffress. *Neuron* **21**, 1235–1238 (1988).
- Grothe, B. & Sanes, D. H. Synaptic inhibition influences the temporal coding properties of medial superior olivary neurons: an in vitro study. *J. Neurosci.* **14**, 1701–1709 (1994).
- Klumpp, R. & Eady, H. Some measurements of interaural time difference thresholds. *J. Acoust. Soc. Am.* **28**, 215–232 (1956).
- Overholt, E., Rubel, E. W. & Hyson, R. L. A circuit for coding interaural time differences in the chick brainstem. *J. Neurosci.* **12**, 1698–1708 (1992).
- Carr, C. E. & Konishi, M. Axonal delay lines for time measurement in the owl's brainstem. *Proc. Natl Acad. Sci. USA* **85**, 8311–8315 (1988).
- Smith, P. H., Joris, P. X. & Yin, T. C. Projections of physiologically characterized spherical bushy cell axons from the cochlear nucleus of the cat: evidence for delay lines to the medial superior olive. *J. Comp. Neurol.* **331**, 245–260 (1993).
- Grothe, B. The evolution of temporal processing in the medial superior olive, an auditory brainstem structure. *Prog. Neurobiol.* **61**, 581–610 (2000).
- Yin, T. C. & Chan, J. C. Interaural time sensitivity in medial superior olive of cat. *J. Neurophys.* **64**, 465–488 (1990).
- Spitzer, M. W. & Semple, M. N. Neurons sensitive to interaural phase disparity in gerbil superior olive: diverse monaural and temporal response properties. *J. Neurophys.* **73**, 1668–1690 (1995).
- Cant, N. B. & Hyson, R. L. Projections from the lateral nucleus of the trapezoid body to the medial superior olivary nucleus in the gerbil. *Hearing Res.* **58**, 26–34 (1992).
- Kuwabara, N. & Zook, J. M. Projections to the medial superior olive from the medial and lateral nuclei of the trapezoid body in rodents and bats. *J. Comp. Neurol.* **324**, 522–538 (1992).
- Grothe, B. & Sanes, D. H. Bilateral inhibition by glycinergic afferents in the medial superior olive. *J. Neurophys.* **69**, 1192–1196 (1993).
- Smith, A. J., Owens, S. & Forsythe, I. D. Characterisation of inhibitory and excitatory postsynaptic currents of the rat medial superior olive. *J. Physiol.* **529**, 681–698 (2000).
- Grothe, B. & Park, T. J. Sensitivity to interaural time differences in the medial superior olive of a small mammal, the Mexican free-tailed bat. *J. Neurosci.* **18**, 6608–6622 (1998).
- Heffner, R. S. & Heffner, H. E. Sound localization and use of binaural cues by the gerbil (*Meriones unguiculatus*). *Behav. Neurosci.* **102**, 422–428 (1988).
- McAlpine, D., Jiang, D. & Palmer, A. R. A neural code for low-frequency sound localization in mammals. *Nature Neurosci.* **4**, 396–401 (2001).
- Grothe, B., Park, T. J. & Schuller, G. Medial superior olive in the free-tailed bat: response to pure tones and amplitude-modulated tones. *J. Neurophys.* **77**, 1553–1565 (1997).
- Rothman, J. S., Young, E. D. & Manis, P. B. Convergence of auditory nerve fibers onto bushy cells in the ventral cochlear nucleus: implications of a computational model. *J. Neurophys.* **70**, 2562–2583 (1993).
- Smith, P. H., Joris, P. X. & Yin, T. C. Anatomy and physiology of principal cells of the medial nucleus of the trapezoid body (MNTB) of the cat. *J. Neurophys.* **79**, 3127–3142 (1998).
- Taschenberger, H. & von Gersdorff, H. Fine-tuning an auditory synapse for speed and fidelity: developmental changes in presynaptic waveform, EPSC kinetics, and synaptic plasticity. *J. Neurosci.* **20**, 9162–9173 (2000).
- Grothe, B. Interaction of excitation and inhibition in processing of pure tone and amplitude-modulated stimuli in the medial superior olive of the mustached bat. *J. Neurophys.* **71**, 706–721 (1994).
- Kapfer, C., Seidl, A. H., Schweizer, H. & Grothe, B. Experience-dependent refinement of inhibitory inputs to auditory coincidence-detector neurons. *Nature Neurosci.* **5**, 257–253 (2002).
- Willott, J. F., Milbrandt, J. C., Bross, L. S. & Caspary, D. M. Glycine immunoreactivity and receptor binding in the cochlear nucleus of c57bl/6j and cba/caj mice: effects of cochlear impairment and aging. *J. Comp. Neurol.* **385**, 405–414 (1997).
- Goldberg, J. M. & Brown, P. B. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *J. Neurophys.* **32**, 613–636 (1969).
- Havey, D. C. & Caspary, D. M. A simple technique for constructing 'piggy-back' multibarrel microelectrodes. *Electroencephalogr. Clin. Neurophys.* **48**, 249–251 (1980).
- Brand, A., Urban, A. & Grothe, B. Duration tuning in the mouse auditory midbrain. *J. Neurophys.* **84**, 1790–1799 (2000).
- Yin, T. C. & Kuwada, S. Binaural interaction in low-frequency neurons in inferior colliculus of the cat. III. Effects of changing frequency. *J. Neurophys.* **50**, 1020–1042 (1983).
- Brughera, A. R., Stutman, E. R., Carney, L. H. & Colburn, H. S. A model with excitation and inhibition for cells in the medial superior olive. *Audit. Neurosci.* **2**, 219–233 (1996).
- Carney, L. H. A model for the responses of low-frequency auditory-nerve fibers in cat. *J. Acoust. Soc. Am.* **93**, 401–417 (1993).

Acknowledgements

We thank G. Breutel for technical assistance, C. Kapfer for suggestions, and M. Goetz for critical comments on the manuscript. This work was supported by the German Research Foundation (A.B., O.B., B.G.) and by the Medical Research Council (T.M., D.M.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.G. (e-mail: bgrothe@neuro.mpg.de).

Nogo-66 receptor antagonist peptide promotes axonal regeneration

Tadzia GrandPré*, Shuxin Li* & Stephen M. Strittmatter

Department of Neurology and Section of Neurobiology, Yale University School of Medicine, P.O. Box 208018, New Haven, Connecticut 06520, USA
* These authors contributed equally to this work

Myelin-derived axon outgrowth inhibitors, such as Nogo, may account for the lack of axonal regeneration in the central nervous system (CNS) after trauma in adult mammals. A 66-residue domain of Nogo (Nogo-66) is expressed on the surface of oligodendrocytes¹ and can inhibit axonal outgrowth through an axonal Nogo-66 receptor (NgR)². The IN-1 monoclonal antibody recognizes Nogo-A and promotes corticospinal tract regeneration and locomotor recovery^{3–5}; however, the undefined nature of the IN-1 epitope in Nogo, the limited specificity of IN-1 for Nogo, and nonspecific anti-myelin effects have prevented a firm conclusion about the role of Nogo-66 or NgR. Here, we identify competitive antagonists of NgR derived from amino-terminal peptide fragments of Nogo-66. The Nogo-66(1–40) antagonist peptide (NEP1–40) blocks Nogo-66 or CNS myelin inhibition of axonal outgrowth *in vitro*, demonstrating that NgR mediates a significant portion of axonal outgrowth inhibition by myelin. Intrathecal administration of NEP1–40 to rats with mid-thoracic spinal cord hemisection results in significant axon growth of the corticospinal tract, and improves functional recovery. Thus, Nogo-66 and NgR have central roles in limiting axonal regeneration after CNS injury, and NEP1–40 provides a potential therapeutic agent.

Previously, we sought to determine which residues within the Nogo-66 domain are responsible for inhibition of axon outgrowth¹. Five 25-residue peptides, consisting of overlapping segments of the Nogo-66 sequence, were assessed in axonal outgrowth assays. Only the peptide consisting of residues 31–55 is inhibitory, demonstrating the role of these residues in receptor activation. However, the inhibitory potency of this peptide is three orders of magnitude less than the entire 66-residue fragment, suggesting that regions other than residues 31–55 contribute to high-affinity binding of Nogo-66 to its receptor. To determine which Nogo-66 residues are required for high-affinity binding, we produced fusion proteins of placental alkaline phosphatase (AP) and various fragments of Nogo-66. Several of the Nogo-66 deletion mutants bind as avidly as Nogo-66 to COS7 cells expressing NgR (Fig. 1a). Notably, the deletion of residues 34–66, encompassing half of the Nogo-66 sequence, does not prevent high-affinity binding by the Nogo-66(1–33) fusion protein. Binding of the Nogo-66(1–31) fusion protein is slightly reduced (dissociation constant, $K_d = 24 \pm 11$ nM; Fig. 1a and data not shown) compared with the Nogo-66(1–33) and longer fusion proteins (K_d values = $11–14 \pm 3$ nM). Additional deletions from the carboxyl end of Nogo-66 markedly reduce NgR binding, such that proteins containing fewer than the 31 N-terminal residues of Nogo-66 exhibit no binding to NgR under these conditions. The necessity of residues at the N terminus was tested in the setting of the Nogo-66(1–40) fusion protein. Deletion of five N-terminal residues in the Nogo-66(6–40) protein significantly reduces binding, and deletion of ten residues (Nogo-66(11–40) protein) abolishes binding. Thus, residues 1–31 seem to be sufficient for high-affinity binding to NgR. Within the N-terminal portion of Nogo-66, at least some of the first ten residues (RIYKGVQA) and residues in the 30–33 region (SEEL) are necessary for maximal binding. The necessity of specific residues from 11–30 for high-affinity binding is not tested here.

As the minimal high-affinity binding domain has little overlap with the 31–55-residue peptide, which exhibits weak agonist activity, we considered the possibility that the N-terminal segment of Nogo-66 might bind to, but not activate, NgR in neurons. Although the full AP–Nogo-66 domain is a potent growth-cone-collapsing agent, the truncated AP–Nogo fusion proteins do not induce growth-cone collapse in embryonic day 12 (E12) chick dorsal root ganglion (DRG) explant cultures (Fig. 1b). Thus, high-affinity binding to NgR can be dissociated from activation of an inhibitory signal through NgR.

The characteristics of the C-terminal-deleted Nogo-66 derivatives suggest that they might function competitively as NgR antagonists. A synthetic Nogo-66(1–40) peptide (Nogo extracellular peptide, residues 1–40, NEP1–40) does not collapse E12 chick DRG growth cones, even at 1 μ M concentrations, consistent with the fusion protein results (Fig. 1d, e). To test for receptor antagonist activity, NEP1–40 was added to E12 chick DRG explant cultures in conjunction with the NgR agonists glutathione *S*-transferase (GST)–Nogo-66 and AP–Nogo-66. NEP1–40 blocks both AP–Nogo-66 binding (Fig. 1c) and GST–Nogo-66-induced growth-cone collapse (Fig. 1d), as expected for a competitive NgR antagonist. This effect

is specific for NgR in that growth-cone collapse by semaphorin 3A (Sema3A) and phorbol ester are not altered by NEP1–40 (Fig. 1e and data not shown). We also assessed the ability of NEP1–40 to neutralize Nogo-66-induced inhibition of neurite outgrowth (Fig. 1f). Soluble NEP1–40 alone has no observable effect on neurite outgrowth from these cells, but it reverses the inhibitory effect of bound GST–Nogo-66. A scrambled sequence version of NEP1–40 does not possess NgR antagonist activity. A second domain of Nogo-A, Amino-Nogo, seems to inhibit outgrowth by a mechanism independent of NgR^{2,6}, and is not blocked by NEP1–40 (Fig. 1f). Taken together, these studies demonstrate that NEP1–40 is a potent NgR competitive antagonist *in vitro*.

The identification of NEP1–40 as a selective NgR antagonist allows an assessment of the relative contribution of NgR ligands to the inhibitory quality of CNS myelin on axon outgrowth. When applied to E12 chick DRG explant cultures, purified CNS myelin potently collapses growth cones and inhibits axonal extension. The presence of NEP1–40 attenuates outgrowth inhibition by CNS myelin, demonstrating that NgR is a significant mediator of inhibition under these conditions (Fig. 2). The potency of NEP1–40 in reversing myelin inhibition is characterized by an effector concentration for half-maximum response (EC_{50}) of 50 nM peptide (Fig. 2b), quite similar to the K_d of AP–Nogo-66(1–40) for NgR (Fig. 1). Myelin inhibition of axon growth is not completely blocked by NEP1–40 even at high peptide concentrations (Fig. 2c). This NEP1–40-insensitive axon outgrowth inhibition by CNS myelin is probably attributable to molecules other than Nogo-66, such as

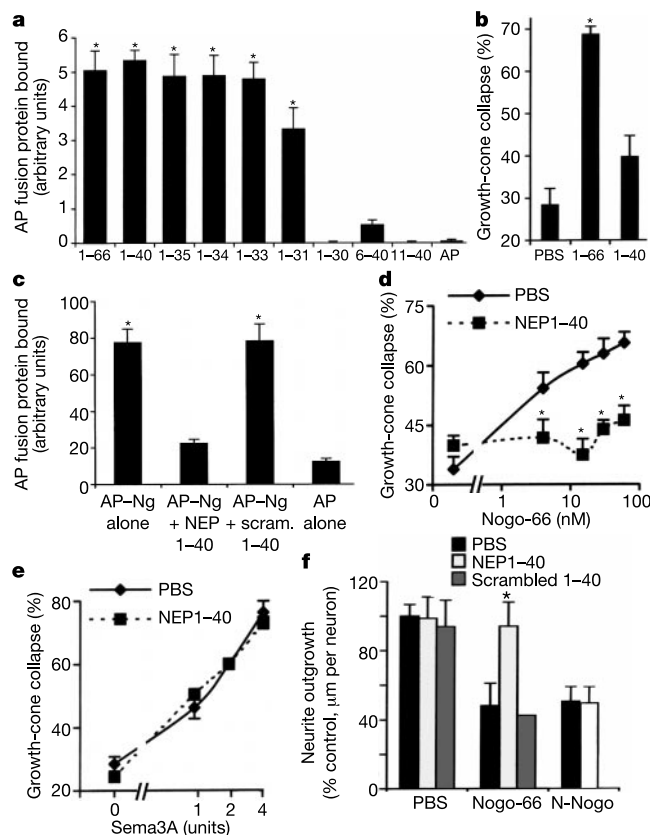


Figure 1 NEP1–40 is a competitive antagonist of Nogo-66. **a**, Conditioned medium containing the indicated Nogo-66-derived AP fusion proteins (28 nM) was applied to COS7 cells expressing NgR. **b**, E12 chick DRG growth-cone morphology after a 30-min exposure to 20 nM AP fusion protein. **c**, Binding of 5 nM AP–Nogo-66 or 5 nM AP to E12 chick DRG axons quantified in the presence or absence of 1 μ M NEP1–40 or scrambled NEP1–40. Ng, Nogo. **d**, **e**, Growth-cone collapse measured in cultures pre-treated with either 1 μ M NEP1–40 or buffer before a 30-min exposure to various concentrations of GST–Nogo-66 (**d**), or Sema3A (**e**). **f**, Neurite outgrowth is reported for dissociated E12 chick DRG neurons grown for 5–7 h on bound substrate coating (GST–Nogo-66 (7.5 ng), Amino-Nogo (N-Nogo; 7.5 ng) or PBS) after treatment with 1 μ M NEP1–40, a scrambled version of NEP1–40, or buffer. Data are the means $\pm$ s.e.m. from 5–6 determinations. Those values statistically different from control are indicated with an asterisk ($P < 0.05$, Student's *t*-test).

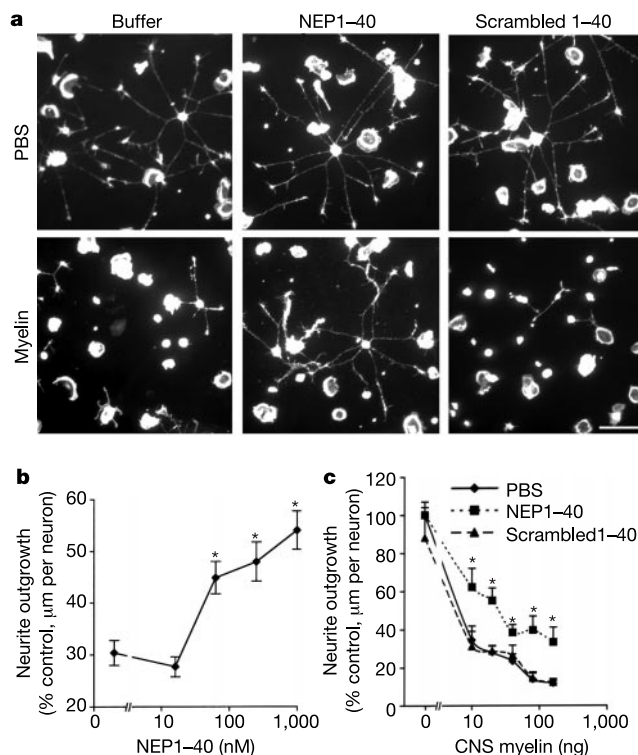


Figure 2 NEP1–40 partially blocks inhibitory activity of CNS myelin. **a**, Dissociated E12 chick DRG cultures grown on a bound substrate coating (CNS myelin or PBS) with 1 μ M soluble NEP1–40, a scrambled version of NEP1–40, or buffer. Scale bar, 75 μ m. **b**, E12 chick DRG neurons were cultured on a substrate coated with 20 ng myelin protein per 3- μ l spot in the presence of the indicated concentrations of NEP1–40. Outgrowth in the absence of CNS myelin is 100%. **c**, Neurite outgrowth from E12 chick dissociated DRGs as in **a** was measured. Cultures were grown 5–7 h on a bound substrate coating (CNS myelin or PBS) with soluble NEP1–40, scrambled NEP1–40, or buffer. Means $\pm$ s.e.m. from 4–9 experiments are reported. The NEP1–40 values are statistically different from the corresponding control (asterisk, $P < 0.05$, Student's *t*-test).

Amino-Nogo (Fig. 1)², myelin-associated glycoprotein (MAG)^{7,8} or chondroitin sulphate proteoglycans⁹.

To the extent that NgR is responsible for limiting axonal regeneration *in vivo* after trauma, NgR antagonism by the NEP1-40 peptide should stimulate regeneration after CNS injury. We delivered the peptide or vehicle intrathecally to rats at the site of a mid-thoracic dorsal hemisection injury through an osmotic minipump. A dose of 75 µg NEP1-40 per kg body weight per day was administered for four weeks in this protocol. The integrity of the descending corticospinal tract (CST) was traced by injection of biotin dextran amine (BDA) into the motor cortex. In rats receiving the vehicle control, the prominent dorsal CST (DCST) is tightly bundled above the lesion site (Fig. 3) and stops abruptly at the transection site. Fewer fibres are observed in the ventral CST (VCST) and only a few dorsolateral CST fibres are detectable (Fig. 3a–c). This pattern is consistent with published reports¹⁰. Below the level of the lesion, a small number of fibres are present but these are confined to the position of the uninjured VCST (Fig. 4e–g).

Samples from injured rats treated with NEP1-40 exhibit an entirely different pattern of labelling. Rostral to the lesion, numerous ectopic fibres sprouting from the labelled CST are observed (Fig. 3b, c). In some cases, projections cross from the DCST near the midline to the circumference of the cord, becoming intermingled with the dorsolateral CST. The sprouting axons extend through grey matter to a greater extent than white matter. A measure of ectopic sprouting adjacent to the DCST reveals a greater than tenfold increase in the peptide-treated animals (Fig. 3b, c). Multiple fibres of this type were seen in every injured animal that had been treated

with the peptide. This cannot be attributed to different degrees of BDA uptake, as the total number of fibres in the dorsal and ventral CSTs above the lesion is indistinguishable between peptide and control sections (Fig. 3a). The sprouting response is due to a combination of NEP1-40 plus axotomy, as no such change is observed in the CST of uninjured rats treated with intrathecal NEP1-40 (Fig. 3b). Similarly, no collateral sprouts from the uninjured VCST of hemisectioned rats treated with NEP1-40 are observed rostral to the lesion (data not shown).

Longitudinal sections across the lesion site in both groups of rats reveal scarring and cavity formation extending from the dorsal surface to, or slightly beyond, the central canal (Fig. 3d–f). The linear projection of the DCST ends at or immediately above the injury site in all animals. In sections from the control group of animals, only occasional fibres in the VCST extend rostral to caudal in a linear fashion, consistent with the labelling of the uninjured VCST. In the NEP1-40 group, a plexus of convoluted fibres crosses from the rostral to caudal cord primarily within the ventral half of the cord (Fig. 3e, f).

Multiple ectopic CST fibres are present throughout the grey and white matter of the lower thoracic spinal cord of the NEP1-40-treated rats (Fig. 4b, d). Even at distances of 11–15 mm caudal to the lesion site, 20–30 CST fibres are observed in each section from these animals, as compared with 0–2 fibres in vehicle-treated rats (Fig. 4e, f). Most of these fibres do not possess the linear trajectory of uninjured fibres but follow branching and tortuous courses. A significant fraction of the fibres are located in the dorsal half of the cord, and some are present in the white matter of the dorsal

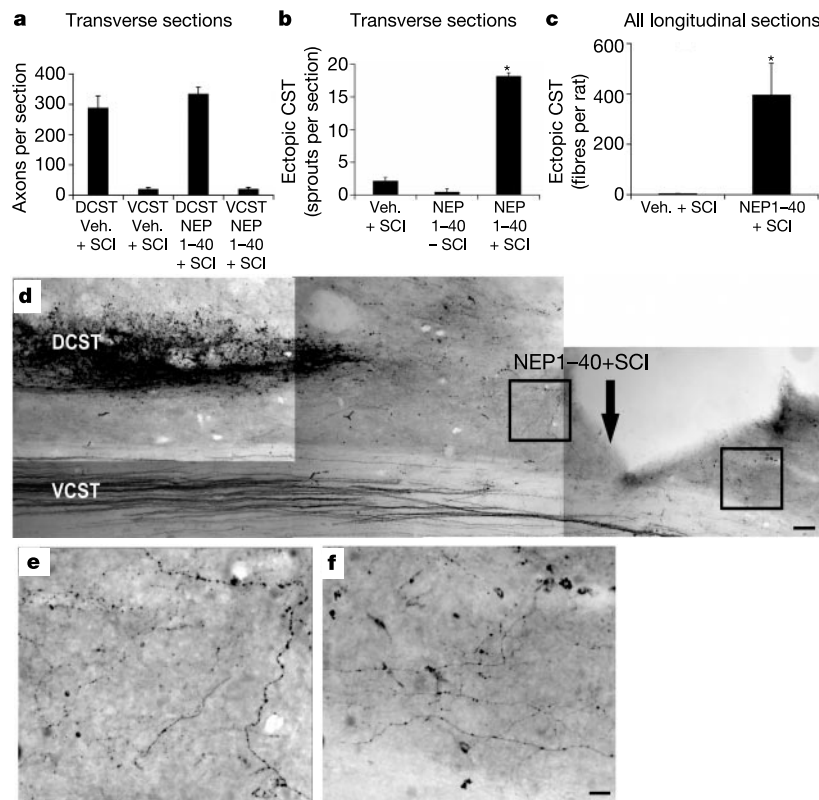


Figure 3 NEP1-40 promotes dorsal CST sprouting rostral to mid-thoracic dorsal hemisection. **a**, Number of BDA-labelled fibres within the DCST or VCST from vehicle (Veh.) and NEP1-40-treated rats in transverse sections 11–15 mm above the lesion. **b**, Number of BDA-labelled fibres outside of the DCST and longer than 200 µm in the transverse plane from transverse sections. **c**, The total number of ectopic fibres in all consecutive sagittal sections lateral to the DCST and 1–10 mm rostral to the lesion site. Only those fibres with a dorsoventral projection of at least 200 µm are scored in order to exclude longitudinally projecting fibres. SCI, spinal cord injury. **d**, Parasagittal section

containing the site of the transection (arrow) from a NEP1-40-treated rat demonstrates numerous labelled fibres in the DCST ending at the injury site. Numerous branched, sprouting fibres are seen in the ventral half of the cord in the two boxed areas. Higher magnification views of these areas in **e** (left box) and **f** (right box) illustrate the meandering course of the regenerating fibres in the NEP1-40-treated rats. A few VCST fibres are seen in the lower half of the figure. Scale bar, 100 µm (**d**); 25 µm (**e**, **f**). Means + s.e.m. from six rats in each group are reported. The NEP1-40 values are statistically different from control (asterisk, $P < 0.05$, Student's *t*-test).

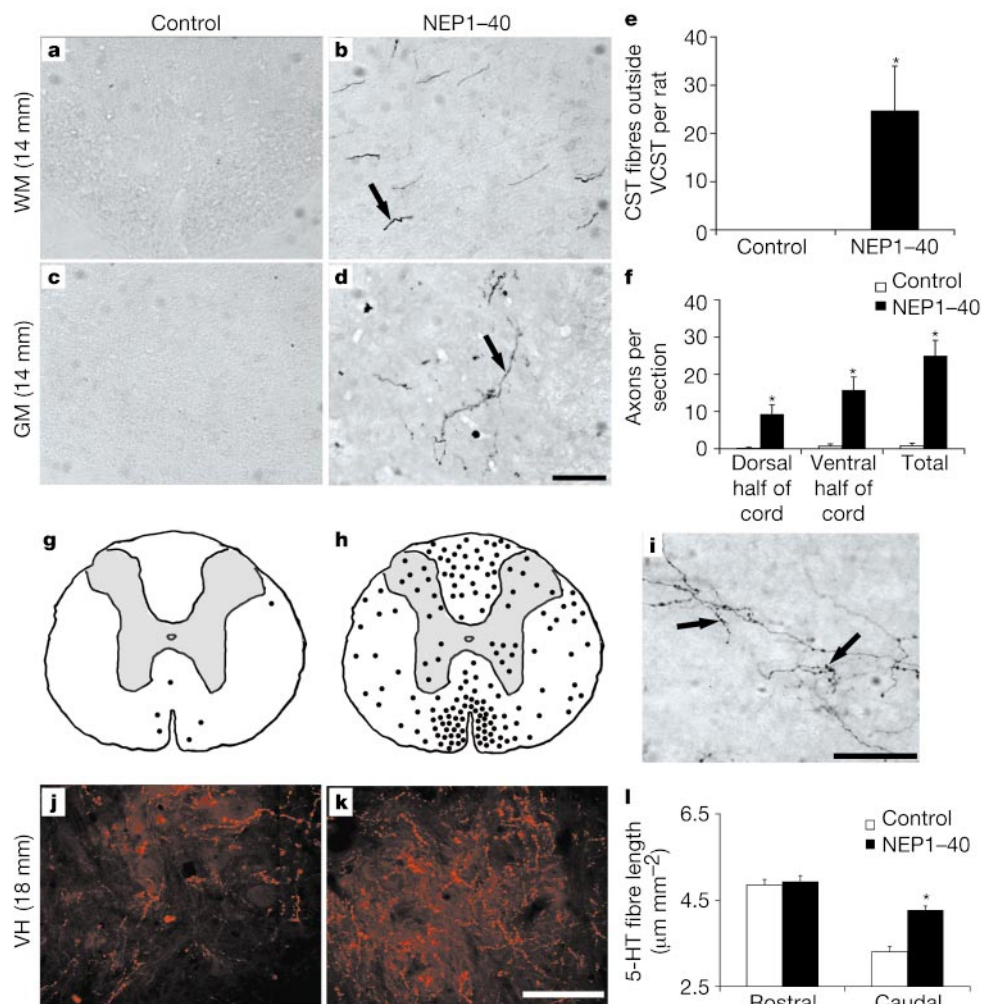


Figure 4 Numerous regenerating CST fibres caudal to a mid-thoracic dorsal hemisection. **a–d**, Transverse sections of dorsal white matter (WM; **a**, **b**) or dorsal grey matter (GM; **c**, **d**) at a level 14 mm caudal to the hemisection reveal no CST fibres in sections from vehicle-treated animals (**a**, **c**), but numerous CST axons (arrows) in the peptide-treated animals (**b**, **d**). **e**, Number of fibres outside of the ventral CST in all consecutive sagittal sections from the tissue of single rats at a level 1–10 mm caudal to the lesion site. **f**, Number of BDA-labelled fibres per section, as in **a–d**, measured in transverse sections 11–15 mm caudal to the lesion. Means $\pm$ s.e.m. from six rats in each group are reported. **g**, **h**, Locations of labelled fibres in transverse sections at a level 11–15 mm below the

lesion site illustrated as a dot for one section from each of five rats (vehicle, **g**; NEP1-40, **h**). The diagram reflects the total number of fibres from five rats. **i**, CST fibres in the ventral horn of an injured rat treated with NEP1-40 exhibit specializations (arrows) consistent with synaptic boutons. **j**, **k**, Serotonin immunoreactivity in the ventral horn (VH) 18 mm below the dorsal hemisection is greater for NEP1-40-treated rats (**k**) than for vehicle-treated animals (**j**). **l**, Serotonin fibre density in the ventral horn was measured 20 mm rostral or 20 mm caudal to a hemisection in control or NEP1-40-treated rats. Scale bars, 25 μ m. Means $\pm$ s.e.m. from six rats in each group are reported. Where indicated, the NEP1-40 values are statistically different from control (asterisk, $P < 0.01$, Student's *t*-test).

column, the site of the injured DCST (Fig. 4h). Several factors make it unlikely that collateral sprouting from uninjured VCST fibres accounts for a significant fraction of the caudal CST fibres in rats with spinal-cord injury treated with NEP1-40. NEP1-40 does not induce collateral sprouting from the DCST in the upper thoracic cord of uninjured rats (Fig. 3b) or the upper thoracic VCST of injured rats. The number of fibres in any one section (approximately 25, Fig. 4e, f) is much greater than the average number in the control VCST caudal to the lesion (about 1, Fig. 4f). Therefore, the numerous regenerating CST fibres in the spinal cord caudal to the injury probably derive from long-distance growth of the numerous ectopic DCST sprouts identified above the lesion site. High magnification views of CST fibres in the lower thoracic ventral horn of rats with spinal cord injury treated with NEP1-40 demonstrate morphology consistent with synaptic boutons, suggesting that functional connections are formed by these regenerating fibres (Fig. 4i).

The dorsal hemisection severs a fraction of serotonin-containing raphespinal fibres. Four weeks after injury, the lower thoracic ventral

horn of control rats exhibits a 40% reduction in the density of 5-hydroxytryptamine (5-HT)-positive fibres compared with sections rostral to the spinal cord injury (Fig. 4j–l). In samples from injured rats treated with NEP1-40, the density of 5-HT-positive fibres from the lower thoracic ventral horn is nearly identical to control values in rostral sections. Thus, NEP1-40 treatment promotes the regeneration of both corticospinal and raphespinal axons. The regeneration of 5-HT fibre density may be due to a combination of collateral sprouting in the caudal spinal cord and long distance axonal growth from severed serotonin-containing axons.

A standardized BBB locomotor score¹¹ (0–21, with 21 being normal function) was used to assess the degree of functional recovery after dorsal hemisection of the spinal cord with or without NEP1-40 treatment (Fig. 5). Those animals receiving NEP1-40 have significantly higher BBB scores by 14 days after spinal cord injury, and this improvement with NEP1-40 persists through 28 days. At the time points of two and seven days, but not at the 4-h time point, there is a lesser degree of improvement in the NEP1-40 group.

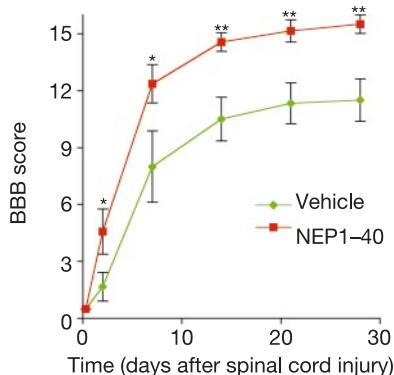


Figure 5 NEP1-40 promotes locomotor recovery after spinal cord injury. The BBB score was determined at the indicated times after mid-thoracic dorsal hemisection in vehicle-treated or NEP1-40-treated rats. The first time point is 4 h after injury when the effects of anaesthetic are nil. Mean $\pm$ s.e.m. from 6 or 7 rats in each group. Where indicated, the NEP1-40 values are statistically different from control for each day (asterisk, $P < 0.05$; double asterisk, $P < 0.01$, Student's t -test).

It may be that functional improvements in the first several days relate to increased collateral sprouting from uninjured axons in the caudal spinal cord, whereas the more significant improvements at later times are dependent on long-distance axonal growth from severed axons.

The principal findings of this study are the identification of a NgR antagonist peptide and the demonstration that a selective NgR antagonist promotes axonal regeneration in rats with spinal cord injury. These data establish a role for Nogo-66 and NgR in limiting axonal regeneration *in vivo*. Analysis of ligand binding to the NgR reveals that a high-affinity binding domain is largely separable from the amino acid residues involved in activating NgR signalling in neurons. As the NEP1-40 peptide is a competitive antagonist of the NgR, its ability to attenuate CNS myelin inhibition of axonal growth demonstrates that the NgR mediates a significant fraction of this inhibition. Local administration of the NgR antagonist peptide to animals after mid-thoracic dorsal spinal cord hemisection elicits a clear axon sprouting response and significant functional recovery. The number of CST fibres at a location over 1 cm caudal to the lesion is about 10% of the total labelled fibres above the lesion. The high potency, selectivity and efficacy of NEP1-40 raises the possibility that a NgR antagonist might be an effective therapeutic agent in clinical conditions characterized by a failure of axonal regeneration, including spinal cord injury, brain trauma, white matter stroke and chronic progressive multiple sclerosis. □

Methods

Nogo-66-derived AP fusion proteins and NEP1-40 synthesis

To generate AP fusion proteins, the relevant regions of human Nogo-A complementary DNA were amplified by polymerase chain reaction and ligated into the pcAP-6 vector as described for AP-Nogo-66 (ref. 2). Binding of AP-Nogo fusion proteins to COS7 cells expressing wild-type NgR and to chick E12 DRG neurons has been described². The NEP1-40 peptide (acetyl-RYKGVQIAIQKSDGHPFRAYLESEVAISEELVQKYSNS-amide) was synthesized using tBOC chemistry at the W. M. Keck Biotechnology Center of Yale University. After preparative reverse phase high-performance liquid chromatography purification, a purity greater than 98% was verified by electrospray ionization spray mass spectrometry. Preparation of GST-Nogo-66 (ref. 1), AP-Nogo-66 (ref. 2), Amino-Nogo², Sema3A (ref. 12) and CNS myelin¹³ were as described.

Growth-cone collapse and neurite outgrowth

E12 chick DRG growth-cone collapse and neurite outgrowth assays have been described¹. For antagonism assays of growth-cone collapse, NEP1-40 was added 10 min before collapsing agent. For assays of neurite outgrowth, plastic chamber slides were coated with 100 μ g ml⁻¹ poly-L-lysine, washed, and dried. Three-microlitre drops of PBS containing GST-Nogo-66, Amino-Nogo or CNS myelin were spotted and dried. Slides were then rinsed and coated with 10 μ g ml⁻¹ laminin before addition of dissociated E12 chick DRGs. We added NEP1-40, or a scrambled version of NEP1-40, at the time of cell plating. We incubated cultures for 5–7 h before assessing neurite outgrowth.

Spinal cord transection and axonal tracing

Female Sprague-Dawley rats (175–250 g) were deeply anaesthetized with ketamine (60 mg per kg) and xylazine (10 mg per kg). Laminectomy was conducted at spinal levels of T6 and T7, and the spinal cord was exposed. The dorsal half of the spinal cord was cut with a pair of microscissors to sever the dorsal parts of the corticospinal tracts, and the depth of lesion (approximately 1.8 mm) was assured by passing the sharp part of a number 11 blade across the dorsal half of the cord. An osmotic minipump (Alzet 2004, ~240 μ l volume, 0.30 μ l h⁻¹, 28 day delivery), which was filled with 97.5% PBS plus 2.5% DMSO or 500 μ M NEP1-40 in 97.5% PBS plus 2.5% DMSO, was sutured to muscles under the skin on the back of the animals. A catheter connected to the outlet of the minipump was inserted into the intrathecal space of the spinal cord at the T7 level through a small hole in the dura. For those rats receiving NEP1-40 without spinal cord injury, the laminectomy and minipump placement were accomplished in the same fashion. Two weeks after lesion, a burr hole was made on each side of the skull overlying the sensorimotor cortex of the lower limbs. The anterograde neuronal tracer biotin dextran amine (BDA, 10% in PBS, 3.5 μ l per cortex) was applied at seven injection sites at a depth of 1.5 mm from the cortical surface on each side. Two weeks after BDA injection, the animals were killed by perfusion with PBS, followed by 4% paraformaldehyde. The spinal cord extending from 10 mm rostral to 10 mm caudal from the lesion site was cut parasagittally (50 μ m) on a vibrating microtome. Transverse sections were collected from the spinal cord 11–16 mm rostral to and 11–16 mm caudal to the injury site. The sections were incubated with avidin-biotin-peroxidase complex and the BDA tracer was visualized by nickel-enhanced diaminobenzidine HRP reaction¹⁴. Some sections were processed for serotonin immunohistochemistry (anti-5-HT antibody; Instar) by indirect immunofluorescence. We processed confocal images using NIH Image software.

Throughout the surgery, post-operative care, behavioural observation and histological analysis, researchers were blind to the identity of the compound in the minipump. A total of 50 rats underwent the full procedure, 23 receiving vehicle and 27 receiving NEP1-40. During the first nine days after spinal cord injury, 30% of the control group and 33% of the NEP1-40 group died or were killed owing to poor general health. Thus, 16 control and 18 NEP1-40-treated animals completed the study. One rat was excluded from the anatomical analysis owing to an incomplete lesion documented by numerous uninjured dorsal CST fibres at four weeks, and a lack of disability (BBB score of 14) at two days after injury. For axon counts, sections from the first six rats processed were scored for each treatment group. In axon counts, a cluster of branched and connected fibres was counted as one; only labelled fibres separated by more than 50 μ m were counted separately. For behavioural analysis, six vehicle-treated and seven NEP1-40-treated rats were compared using the BBB scale¹¹.

Received 26 January; accepted 25 March 2002.

- GrandPre, T., Nakamura, F., Vartanian, T. & Strittmatter, S. M. Identification of the Nogo inhibitor of axon regeneration as a Reticulon protein. *Nature* **403**, 439–444 (2000).
- Fournier, A. E., GrandPre, T. & Strittmatter, S. M. Identification of a receptor mediating Nogo-66 inhibition of axonal regeneration. *Nature* **409**, 341–346 (2001).
- Chen, M. S. *et al.* Nogo-A is a myelin-associated neurite outgrowth inhibitor and an antigen for monoclonal antibody IN-1. *Nature* **403**, 434–439 (2000).
- Bregman, B. S. *et al.* Recovery from spinal cord injury mediated by antibodies to neurite growth inhibitors. *Nature* **378**, 498–501 (1995).
- Merkler, D. *et al.* Locomotor recovery in spinal cord-injured rats treated with an antibody neutralizing the myelin-associated neurite growth inhibitor Nogo-A. *J. Neurosci.* **21**, 3665–3673 (2001).
- Prinjha, R. *et al.* Inhibitor of neurite outgrowth in humans. *Nature* **403**, 383–384 (2000).
- Mukhopadhyay, G., Doherty, P., Walsh, F. S., Crocker, P. R. & Filbin, M. T. A novel role for myelin-associated glycoprotein as an inhibitor of axonal regeneration. *Neuron* **13**, 757–767 (1994).
- McKerracher, L. *et al.* Identification of myelin-associated glycoprotein as a major myelin-derived inhibitor of neurite growth. *Neuron* **13**, 805–811 (1994).
- Niederost, B. P., Zimmermann, D. R., Schwab, M. E. & Bandtlow, C. E. Bovine CNS myelin contains neurite growth-inhibitory activity associated with chondroitin sulfate proteoglycans. *J. Neurosci.* **19**, 8979–8989 (1999).
- Weidner, N., Ner, A., Salimi, N. & Tuszynski, M. H. Spontaneous corticospinal axonal plasticity and functional recovery after adult central nervous system injury. *Proc. Natl Acad. Sci. USA* **98**, 3513–3518 (2001).
- Basso, D. M., Beattie, M. S. & Bresnahan, J. C. A sensitive and reliable locomotor rating scale for open field testing in rats. *J. Neurotrauma* **12**, 1–21 (1995).
- Takahashi, T., Nakamura, F. & Strittmatter, S. M. Neuronal and non-neuronal collapsin-1 binding sites in developing chick are distinct from other semaphorin binding sites. *J. Neurosci.* **17**, 9183–9193 (1997).
- Fournier, A. E. *et al.* Semaphorin3A enhances endocytosis at sites of receptor-F-actin colocalization during growth cone collapse. *J. Cell Biol.* **149**, 411–422 (2000).
- Thallmair, M. *et al.* Neurite growth inhibitors restrict plasticity and functional recovery following corticospinal tract lesions. *Nature Neurosci.* **1**, 124–131 (1998).

Acknowledgements

We thank Y. Fu for technical assistance. This work was supported by grants to S.M.S. from the NIH, the McKnight Foundation for Neuroscience, and Biogen Inc. T.G. is a Bayer Predoctoral Scholar and S.M.S. is an Investigator of the Patrick and Catherine Weldon Donaghy Medical Research Foundation.

Competing interests statement

The authors declare competing financial interests: details accompany the paper on *Nature's* website (<http://www.nature.com>).

Correspondence and requests for materials should be addressed to S.M.S. (e-mail: stephen.strittmatter@yale.edu).

A component of innate immunity prevents bacterial biofilm development

Pradeep K. Singh[†], Matthew R. Parsek[‡], E. Peter Greenberg^{‡§} & Michael J. Welsh^{*§||}

^{*} Department of Internal Medicine, [‡] Department of Microbiology, and ^{||} Department of Physiology and Biophysics and Howard Hughes Medical Institute, University of Iowa College of Medicine, Iowa City, Iowa 52242, USA
[†] Department of Civil Engineering, Northwestern University, Evanston, Illinois 60208, USA
[§] W. M. Keck Foundation Microbial Communities and Cell Signaling Laboratory, Iowa City, Iowa 52242, USA

Antimicrobial factors form one arm of the innate immune system, which protects mucosal surfaces from bacterial infection^{1–3}. These factors can rapidly kill bacteria deposited on mucosal surfaces and prevent acute invasive infections^{1–4}. In many chronic infections, however, bacteria live in biofilms, which are distinct, matrix-encased communities specialized for surface persistence^{5–7}. The transition from a free-living, independent existence to a biofilm lifestyle can be devastating, because biofilms notoriously resist killing by host defence mechanisms and antibiotics^{5,8}. We hypothesized that the innate immune system possesses specific activity to protect against biofilm infections. Here we show that lactoferrin, a ubiquitous and abundant constituent of human external secretions, blocks biofilm development by the opportunistic pathogen *Pseudomonas aeruginosa*. This occurs at lactoferrin concentrations below those that kill or prevent growth. By chelating iron, lactoferrin stimulates twitching, a specialized form of surface motility, causing the bacteria to wander across the surface instead of forming cell clusters and biofilms. These findings reveal a specific anti-biofilm defence mechanism acting at a critical juncture in biofilm development, the time bacteria stop roaming as individuals and aggregate into durable communities.

Their occurrence in chronic infections brands bacterial biofilms a major medical problem^{5,6,8}. The airway infection by *P. aeruginosa* that afflicts people with cystic fibrosis is a prime example of a biofilm infection^{6,7}. Once this infection develops, *P. aeruginosa* colonize the airways for life, causing lung destruction and eventually death⁹. Normal mucosal surfaces resist biofilm infections despite

continual exposure to pathogenic bacteria. Rapid killing of deposited organisms probably accounts for some of this resistance¹. At times, however, bacteria remain on mucosal surfaces for prolonged periods, for example, during infections such as bronchitis, conjunctivitis, or infections associated with foreign bodies. This led us to hypothesize that mucosal surfaces might possess an anti-biofilm defence, perhaps using antimicrobial factors, a phylogenetically conserved limb of the innate immune system¹⁰.

To test this hypothesis, we asked whether human lactoferrin could inhibit biofilm development in *P. aeruginosa*. We investigated the effect of lactoferrin because it is ubiquitous, and it is among the most abundant proteins present in surface secretions. Substantial concentrations of iron-unsaturated lactoferrin¹¹ are found in tears (1–4 mg ml^{−1})¹² and airway secretions (~0.4–1.0 mg ml^{−1})^{13,14}. In addition, breast milk delivers large amounts (~3–7 mg ml^{−1})¹⁵ to the undeveloped digestive systems of babies. At high concentrations, lactoferrin is known to limit bacterial growth by sequestering iron. In this regard it acts like other nutrient-depriving host defence molecules, such as transcobalamins (which bind vitamin B12)¹⁶ and calprotectin (which binds zinc)¹⁷. Lactoferrin can also be bactericidal by binding lipopolysaccharide and disrupting bacterial membranes, and it can enhance killing by other antibiotics^{18,19}. To determine whether lactoferrin has anti-biofilm activity that is distinct from these known properties, we examined the effect of a subinhibitory concentration of lactoferrin (20 µg ml^{−1}) on biofilm

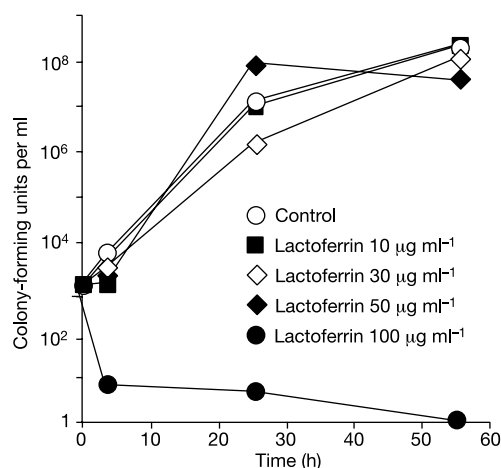


Figure 1 Growth of *P. aeruginosa* in the presence of lactoferrin. Results are representative of three experiments.

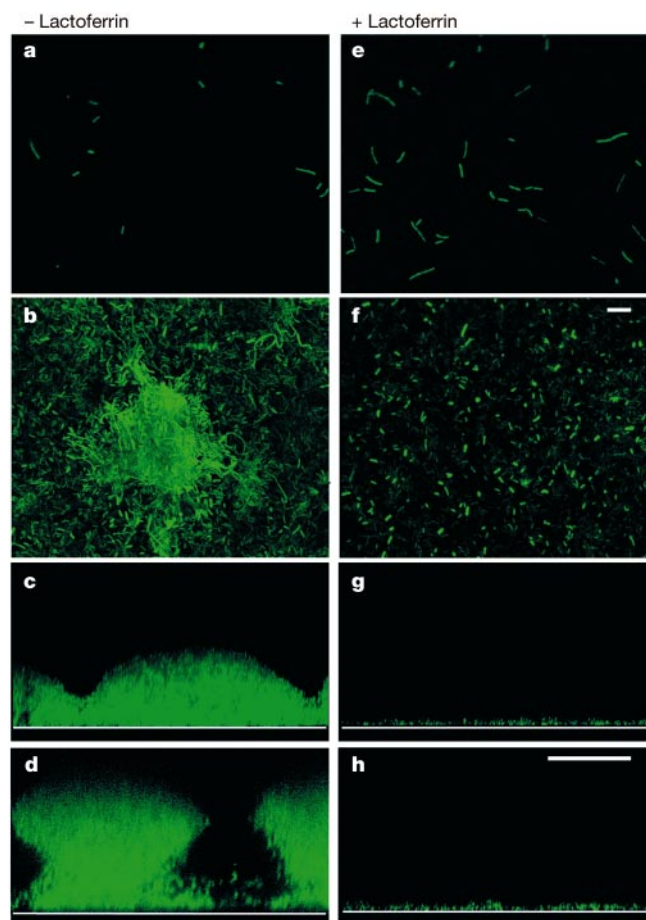


Figure 2 Confocal microscopic images of GFP-labelled *P. aeruginosa* in biofilm flow cells perfused with lactoferrin-free (a–d) and lactoferrin-containing (20 µg ml^{−1}) (e–h) media. Images were obtained 4 h (a, e), 24 h (b, f), 3 days (c, g) and 7 days (d, h) after inoculating the flow cells. Images a, b, e and f are top views (x–y plane); scale bar, 10 µm. Images c, d, g and h are side views (x–z plane); scale bar, 50 µm. Results are representative of six experiments.

development. This concentration of lactoferrin did not affect the growth rate of free-swimming *P. aeruginosa* strain PAO1 (Fig. 1).

To evaluate the effect of lactoferrin on biofilm formation, we grew *P. aeruginosa* expressing green fluorescent protein (GFP) in continuous-culture-flow cells and followed biofilm development over time. Flow cell chambers were continuously perfused with biofilm medium with or without lactoferrin. In medium without lactoferrin (Fig. 2a–d), we observed the typical stages of biofilm development^{20,21}. Initially, bacteria attached to the surface (Fig. 2a). Microcolonies were evident after 24 h (Fig. 2b). After 3 days, the microcolonies had enlarged (Fig. 2c). By day 7, towering pillar and mushroom-shaped biofilms had developed (Fig. 2d). Lactoferrin disrupted this pattern of development (Fig. 2e–h). Attached bacteria (Fig. 2e) multiplied, but they failed to form microcolonies (Fig. 2f). Even after prolonged incubation, the bacteria did not assemble into differentiated biofilm structures; in the presence of lactoferrin they remained in a thin layer (Fig. 2g, h). In contrast, exposing mature, 5-day-old biofilms to lactoferrin-containing medium for 48 h failed to alter their structure (not shown). Thus, once they had developed, biofilms were resistant to lactoferrin.

Because lactoferrin prevented biofilm development, we performed additional studies to confirm that the low concentration of lactoferrin did not prevent growth. To show that lactoferrin-containing medium in flow cells could support growth of *P. aeruginosa*, we cultured bacteria in the effluent from a biofilm chamber. *P. aeruginosa* doubled six times in 22 h in this conditioned medium (not shown), verifying that even spent lactoferrin-treated medium did not limit growth. Second, we measured the dividing times of attached bacteria in flow cells using time-lapse video microscopy. Lactoferrin increased the dividing time of attached

cells by 27% (93 min without lactoferrin compared with 127 min with $20 \mu\text{g ml}^{-1}$ lactoferrin). Although this reduced growth rate could decrease the size of microcolonies and biofilms, it could not account for the complete absence of biofilm structure induced by lactoferrin.

Although the time-lapse microscopy showed only small differences in dividing times, it revealed that lactoferrin markedly altered bacterial movement. These differences are represented in Fig. 3a and b, tracing the movement of representative bacteria over the surface of a flow cell. In the absence of lactoferrin (Fig. 3a), the parent bacterium moved across the field of view. When the parent cell divided, the two daughter cells remained near the point of parent cell division. When a daughter cell divided, its progeny also remained near the point of the original cell division. Thus, a microcolony began to form. In the presence of lactoferrin (Fig. 3b), the parental cell also moved across the field of view, and divided into two daughter cells. With lactoferrin, however, the daughter cells moved away from the point of cell division. When one of the daughter cells divided, its progeny also left the site of cell division (time-lapse microscopy images are available as Supplementary Information).

To analyse the changes quantitatively, we defined three behaviours and classified the actions of 40 parental cells and their offspring over three generations. Bacteria that remained stationary from the time they were created by cell division to the time they themselves divided were called squatters. Bacteria that moved away from the division site were called ramblers, and cells that detached from the surface and were swept away by the flow of medium were called flyers. The relative proportions of bacteria engaged in these different behaviours are shown in Fig. 3c and d. In the absence of lactoferrin, most cells were squatters, fewer cells were flyers, and rambling was rare. In the presence of lactoferrin, a significantly larger proportion of cells exhibited rambling behaviour and fewer were squatters. In both cases, the predominant behaviour (squatting without lactoferrin, and rambling with lactoferrin) became more

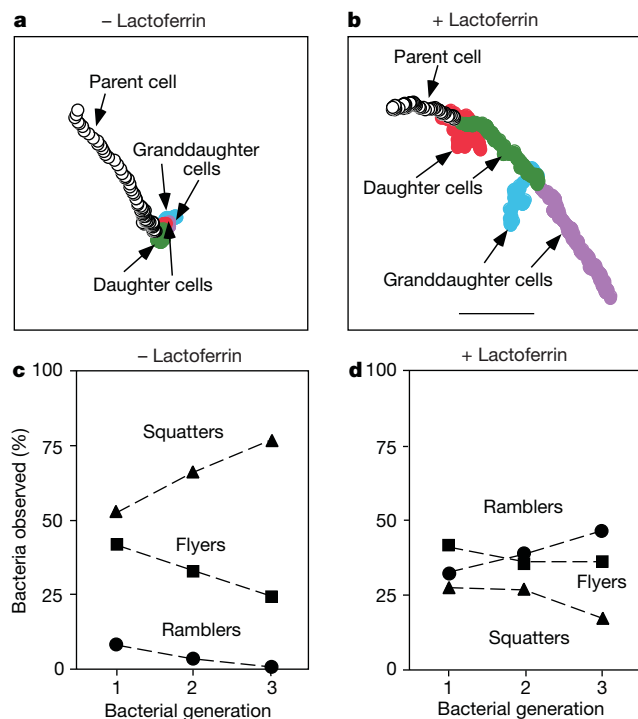


Figure 3 Representations of bacterial behaviours. **a, b**, Movement of representative cells without (**a**) and with (**b**) lactoferrin. Each point represents bacterial position at 1-min intervals. Scale bar, $20 \mu\text{m}$. **c, d**, Effect of lactoferrin on bacterial behaviour without (**c**) and with (**d**) $20 \mu\text{g ml}^{-1}$ lactoferrin. Three behaviours were defined: squatting, flying and rambling, and the proportion of bacteria engaged in each behaviour was assessed over three generations. Data were collected from six different movies and represent the behaviour of a total of 462 bacteria. Proportions of squatters and ramblers were different in control and lactoferrin media ($P < 0.001$, binomial distribution test).

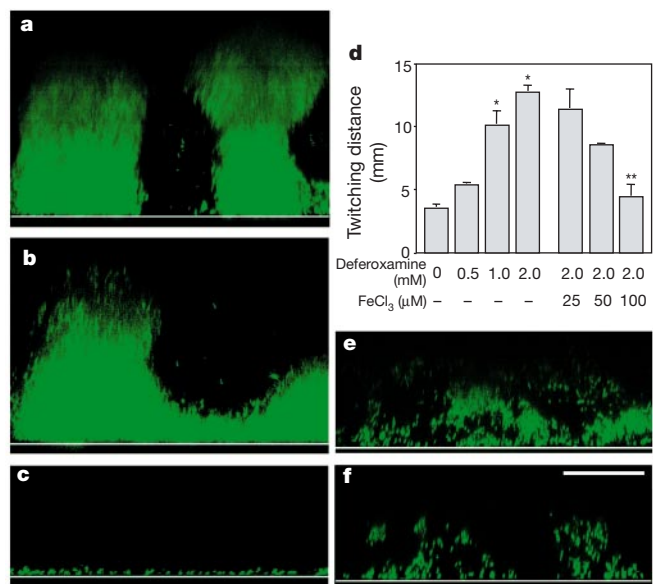


Figure 4 Role of iron in twitching motility and biofilm development. **a–c**, Effect of iron-saturated lactoferrin on biofilm formation in wild-type *P. aeruginosa*. **a**, No lactoferrin; **b**, Fe-saturated lactoferrin; **c**, Fe-unsaturated lactoferrin. **d**, Effect of iron chelation by deferoxamine on *P. aeruginosa* twitching motility. Asterisk, $P < 0.001$, one-way ANOVA, as compared with no deferoxamine. Double asterisk, $P < 0.005$, one-way ANOVA, as compared with 2 mM deferoxamine without FeCl_3 . **e, f**, Effect of control (**e**) and lactoferrin-containing (**f**) medium on biofilm formation by a non-twitching mutant. Scale bar, $50 \mu\text{m}$. Similar results were obtained in three other experiments.

prevalent in subsequent bacterial generations.

We next investigated the mechanism of lactoferrin's action on biofilm development. To examine the role of iron, we compared the activity of iron-saturated lactoferrin to iron-unsaturated lactoferrin (Fig. 4a–c). Unlike iron-unsaturated lactoferrin, iron-saturated lactoferrin did not prevent biofilm formation by *P. aeruginosa*. Conalbumin, a lactoferrin-like host defence protein from chicken eggs, functioned similarly; it prevented biofilm formation in the iron-unsaturated state, but not when saturated with iron (not shown). We also studied deferoxamine, an iron chelator unrelated to lactoferrin that has a lower iron-binding affinity than lactoferrin^{22,23}. At subinhibitory concentrations, it prevented biofilm formation, and time-lapse studies showed that deferoxamine also stimulated bacterial surface motility (not shown). However, with deferoxamine, microcolonies did occasionally form, probably owing to its lower iron-binding affinity. Together, these results suggest that lactoferrin blocks biofilm formation in *P. aeruginosa* by sequestering free iron.

We hypothesized that the increased surface motility induced by iron chelation was due to twitching, a specialized form of surface locomotion mediated by type 4 pili²⁴. To test this, we performed twitching motility assays in which *P. aeruginosa* was inoculated at a point on the bottom of agar plates, and the rate at which bacteria spread over the agar–plastic interface was measured. Deferoxamine stimulated twitching motility in a dose-dependent manner, and this response was blocked by adding iron (Fig. 4d). Thus, as free iron levels decreased, twitching motility increased.

To test further whether lactoferrin prevented biofilm development by stimulating twitching motility, we examined its effect on a *P. aeruginosa* twitching mutant. We reasoned that the mutant would form biofilms in the presence of lactoferrin. The mutant formed microcolonies and irregularly shaped biofilms in both the absence and presence of lactoferrin (Fig. 4e, f), although the biofilms formed in the presence of lactoferrin were somewhat smaller. This stands in contrast to the twitching wild-type strain, where differentiated biofilm formation was completely blocked by lactoferrin (Fig. 4a, c). Taken together, these results indicate that lactoferrin prevents biofilm formation by stimulating bacterial twitching motility. Furthermore, the concentration of lactoferrin that had this effect did not limit the growth of free-swimming bacteria and only slightly reduced the growth of attached cells. Once bacteria were living in an established biofilm, they lost sensitivity to lactoferrin.

Biofilm bacteria are extraordinarily resistant to killing by antimicrobial agents^{5,6,8}. We hypothesized that the surface-attached bacterial layers that formed in the presence of lactoferrin would be less resistant than differentiated biofilms that formed in the absence of lactoferrin. To test this, we grew bacteria in a biofilm reactor on small removable discs. This allowed the antimicrobial susceptibility of the bacterial community to be tested with its

multicellular structure intact. Bacteria were grown with or without conalbumin. Conalbumin was used in these studies because the cost of lactoferrin was prohibitive for the large volume of medium required, and, as described above, lactoferrin and conalbumin affected biofilm formation similarly. After 48 h, the discs were removed from the reactor (and from the conalbumin), and two agents were tested: H₂O₂, which neutrophils use in the oxidative killing of bacteria, and tobramycin, an antibiotic used clinically to treat *P. aeruginosa* infections. Control biofilms were resistant to both agents: 1,000 µg ml⁻¹ tobramycin and 500 mM H₂O₂ had minimal effects on viability after 4 h of treatment (Fig. 5). In contrast, growth in conalbumin decreased resistance to both agents in a dose-dependent manner. Thus, in addition to inhibiting structural differentiation, iron chelation limited the development of an important functional consequence of biofilm formation, antimicrobial resistance.

These findings can be viewed from two perspectives. For the host, the development of a biofilm infection on a normally sterile mucosal surface can have disastrous consequences. Our data suggest that lactoferrin has a previously unrecognized role in host defence. In addition to its well-known bactericidal and bacteriostatic actions, it blocks the formation of *P. aeruginosa* biofilms at a low concentration, keeping the bacteria more vulnerable to killing. This function may serve as a failsafe mechanism to prevent bacteria that survive initial defences from assuming the intractable biofilm state. Secondary immune responses may then be better able to combat the infecting organisms. As noted above, biofilms form in the airways of cystic fibrosis patients and on other compromised mucosal surfaces^{5–7}. Although our data do not address how this occurs in the presence of lactoferrin, there is evidence that lactoferrin may be inactivated by proteolytic cleavage in the lungs of cystic fibrosis patients²⁵. Other reports have shown that the levels of free iron are increased in airway secretions of cystic fibrosis patients^{26,27}. However, these observations were made in patients with established airway infections, and thus their relevance to the pathogenesis of *Pseudomonas* biofilm infection is unclear.

From the bacterial point of view, biofilms are a growth mode specialized for long-term colonization of surfaces. Our data indicate that a higher level of iron is required for biofilm formation than is needed for growth. If the iron level is acceptable, *P. aeruginosa* is cued to stop moving, form microcolonies, and eventually develop into biofilms. If iron levels are not sufficient, the *P. aeruginosa* cells keep moving. This response may protect the bacteria from constructing complex, durable biofilm structures in locations where iron, a critical nutrient, is in short supply. □

Methods

Bacterial strains, plasmids, and growth conditions

Pseudomonas aeruginosa strain PAO1 containing the GFP plasmid pMRP9-1 (ref. 28) was used for most studies. Where indicated, an isogenic twitching motility mutant (a PAO1 pilHJK deletion mutant from J. Kato) containing pMRP9-1 was used. Biofilm medium consisted of 1% Trypticase Soy Broth (Difco). In the growth experiments, about 10³ bacteria from an overnight culture were added to 5 ml of biofilm medium containing various concentrations of lactoferrin or deferoxamine (Sigma). Cultures were incubated in acid-washed tubes at 37 °C with shaking. Colony-forming units were determined by plate counting. The concentration of lactoferrin that inhibited bacterial growth varied somewhat for different lots of lactoferrin but was never less than 30 µg ml⁻¹. The concentration of deferoxamine that slowed growth was never less than 5 mM. *P. aeruginosa* with pMRP9-1 was also grown in lactoferrin-containing effluent medium from flow cells. After 1 day of bacterial growth, effluent was collected on ice, filter sterilized, and growth was assessed as above.

Biofilm experiments

For studies of biofilm formation, wild-type *P. aeruginosa* PAO1 and the twitching motility mutant were grown in flow cells similar to those described previously²⁸; the size of the flow channel was 5 × 35 × 1 mm. An overnight culture diluted to 10⁷ cells per ml in fresh biofilm medium was used as the inoculum and flow was arrested for 45 min. Flow of biofilm medium with and without 20 µg ml⁻¹ of Fe-unsaturated or Fe-saturated lactoferrin, conalbumin, or 3–5 mM deferoxamine (Sigma) was then initiated at a rate of 170 µl min⁻¹. Images were obtained using a Bio-Rad scanning confocal microscope. Bacterial movement was assessed by using time-lapse images acquired at 1 min intervals.

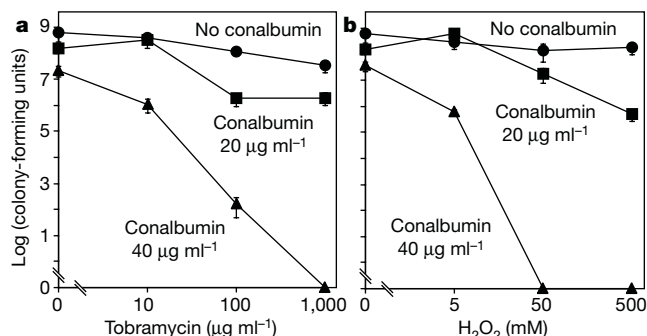


Figure 5 Effect of conalbumin on the antimicrobial susceptibility of *P. aeruginosa* biofilms to tobramycin (a) and H₂O₂ (b). Data are mean ± s.e.m., *n* = 6 from three different experiments.

Motion of bacterial cells was traced visually by following individual cells. VOXblast software (VayTek) was used to obtain the x - y coordinates of bacterial cells. Bacterial behaviors were classified as follows. Squatters remained within a 15 μ m circle drawn around the point of parental cell division until the time of its cell division. Ramblers remained attached to the growth surface, but moved outside the circle. Flyers detached and were carried away by media flow. Dividing times of attached bacteria were measured by counting the number of frames between cell divisions; 60 bacterial divisions were observed and the dividing times were averaged.

Pseudomonas aeruginosa biofilms for susceptibility tests were grown in a rotating disc reactor²⁹. Fe-unsaturated conalbumin (Sigma) was added to standard biofilm medium at indicated concentrations for the duration of biofilm growth. Discs and attached bacteria were then washed three times in distilled water, and treated for 4 h in 1 ml of H₂O₂ (Fisher Scientific) or tobramycin (Eli Lilly) at indicated concentrations. The treated discs were washed three times, and bacteria were removed and dispersed in 2 ml sterile PBS by homogenization (Brinkman Homogenizer). Viable cell numbers were enumerated by plate counting.

Twitching motility assays

Plates for twitching motility assays consisted of biofilm medium plus 1% Noble agar (Difco). Indicated concentrations of deferrioxamine and FeCl₃ (Sigma) were added to molten agar. Plates were dried overnight at room temperature, and *P. aeruginosa* with pMRP9-1 was point inoculated at the bottom of the Petri plate³⁰. After 3 days, the twitching distance along the plastic-agar interface (at the bottom of the agar plate) was measured.

Received 5 December 2001; accepted 14 March 2002.

- Coonrod, J. D. The role of extracellular bactericidal factors in pulmonary host defenses. *Semin. Respir. Infect.* **1**, 118–129 (1986).
- Cole, A. M., Dewan, P. & Ganz, T. Innate antimicrobial activity of nasal secretions. *Infect. Immun.* **67**, 3267–3275 (1999).
- Ganz, T. Antimicrobial proteins and peptides in host defense. *Semin. Respir. Infect.* **16**, 4–10 (2001).
- Nizet, V. *et al.* Innate antimicrobial peptide protects the skin from invasive bacterial infection. *Nature* **414**, 454–457 (2001).
- Costerton, J. W., Stewart, P. S. & Greenberg, E. P. Bacterial biofilms: a common cause of persistent infections. *Science* **284**, 1318–1322 (1999).
- Hoiby, N. *et al.* *Pseudomonas aeruginosa* and the *in vitro* and *in vivo* biofilm mode of growth. *Microbes Infect.* **3**, 23–35 (2001).
- Singh, P. K. *et al.* Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms. *Nature* **407**, 762–764 (2000).
- Xu, K. D., McFeters, G. A. & Stewart, P. S. Biofilm resistance to antimicrobial agents. *Microbiology* **146**, 547–549 (2000).
- Burns, J. L., Ramsey, B. W. & Smith, A. L. Clinical manifestations and treatment of pulmonary infections in cystic fibrosis. *Adv. Pediatr. Infect. Dis.* **8**, 53–66 (1993).
- Hoffmann, J. A., Kafatos, F. C., Janeway, C. A. & Ezekowitz, R. A. Phylogenetic perspectives in innate immunity. *Science* **284**, 1313–1318 (1999).
- Ward, C. G., Bullen, J. J. & Rogers, H. J. Iron and infection: new developments and their implications. *J. Trauma* **41**, 356–364 (1996).
- Abe, T., Nakajima, A., Matsunaga, M., Sakuragi, S. & Komatsu, M. Decreased tear lactoferrin concentration in patients with chronic hepatitis C. *Br. J. Ophthalmol.* **83**, 684–687 (1999).
- Harbitz, O., Jenssen, A. O. & Smidsrod, O. Lysozyme and lactoferrin in sputum from patients with chronic obstructive lung disease. *Eur. J. Respir. Dis.* **65**, 512–520 (1984).
- Thompson, A. B., Bohling, T., Payvandi, F. & Rennard, S. I. Lower respiratory tract lactoferrin and lysozyme arise primarily in the airways and are elevated in association with chronic bronchitis. *J. Lab. Clin. Med.* **115**, 148–158 (1990).
- Hirai, Y. *et al.* Concentrations of lactoferrin and iron in human milk at different stages of lactation. *J. Nutr. Sci. Vitaminol. (Tokyo)* **36**, 531–544 (1990).
- Samson, R. R., Mirtle, C. & McClelland, D. B. Secretory IgA does not enhance the bacteriostatic effects of iron-binding or vitamin B12-binding proteins in human colostrum. *Immunology* **38**, 367–373 (1979).
- Sohnle, P. G., Collins-Lech, C. & Wiessner, J. H. The zinc-reversible antimicrobial activity of neutrophil lysates and abscess fluid supernatants. *J. Infect. Dis.* **164**, 137–142 (1991).
- Ellison, R. T. III The effects of lactoferrin on Gram-negative bacteria. *Adv. Exp. Med. Biol.* **357**, 71–90 (1994).
- Leitch, E. C. & Willcox, M. D. Lactoferrin increases the susceptibility of *S. epidermidis* biofilms to lysozyme and vancomycin. *Curr. Eye Res.* **19**, 12–19 (1999).
- Kolter, R. & Losick, R. One for all and all for one. *Science* **280**, 226–227 (1998).
- Tolker-Nielsen, T. *et al.* Development and dynamics of *Pseudomonas* sp. biofilms. *J. Bacteriol.* **182**, 6482–6489 (2000).
- Keberle, H. The biochemistry of deferrioxamine and its relation to iron metabolism. *Ann. NY Acad. Sci.* **119**, 758–768 (1964).
- Xiao, R. & Kisaalita, W. S. Iron acquisition from transferrin and lactoferrin by *Pseudomonas aeruginosa* pyoverdinin. *Microbiology* **143**, 2509–2515 (1997).
- Martin, P. R., Watson, A. A., McCaul, T. F. & Mattick, J. S. Characterization of a five-gene cluster required for the biogenesis of type 4 fimbriae in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **16**, 497–508 (1995).
- Britigan, B. E., Hayek, M. B., Doebbeling, B. N. & Fick, R. B. J. Transferrin and lactoferrin undergo proteolytic cleavage in the *Pseudomonas aeruginosa*-infected lungs of patients with cystic fibrosis. *Infect. Immun.* **61**, 5049–5055 (1993).
- Stites, S. W., Plautz, M. W., Bailey, K., O'Brien-Ladner, A. R. & Wesseliuss, L. J. Increased concentrations of iron and isoferritins in the lower respiratory tract of patients with stable cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **160**, 796–801 (1999).
- Stites, S. W., Walters, B., O'Brien-Ladner, A. R., Bailey, K. & Wesseliuss, L. J. Increased iron and ferritin content of sputum from patients with cystic fibrosis or chronic bronchitis. *Chest* **114**, 814–819 (1998).

- Davies, D. G. *et al.* The involvement of cell-to-cell signals in the development of a bacterial biofilm. *Science* **280**, 295–298 (1998).
- Hentzer, M. *et al.* Alginate overproduction affects *Pseudomonas aeruginosa* biofilm structure and function. *J. Bacteriol.* **183**, 5395–5401 (2001).
- O'Toole, G. A. & Kolter, R. Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. *Mol. Microbiol.* **30**, 295–304 (1998).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank E. Brutinel, T. Moninger and M. Neville for technical assistance, J. Kato of Hiroshima University for the twitching mutant, A. Berger for advice on data analysis, and our laboratory colleagues for discussions. This research was supported by the Howard Hughes Medical Institute, the National Heart, Lung and Blood Institute, the National Institute of General Medical Science, the Cystic Fibrosis Foundation, and the W. M. Keck Foundation Microbial Communities and Cell Signaling Laboratory. P.K.S. is the recipient of a NIH Mentored Physician Scientist Award, and a Cystic Fibrosis Foundation Leroy Matthews Award. M.J.W. is an Investigator of the Howard Hughes Medical Institute.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.K.S. (e-mail: pradeep-singh@uiowa.edu).

Single and multiple vesicle fusion induce different rates of endocytosis at a central synapse

Jian-Yuan Sun*, Xin-Sheng Wu* & Ling-Gang Wu

Departments of Anesthesiology, Anatomy and Neurobiology, Washington University School of Medicine, St Louis, Missouri 63110, USA

* These authors contributed equally to this work

During synaptic transmission, neurotransmitter-laden vesicles fuse with the presynaptic membrane and discharge their contents into the synaptic cleft. After fusion, the vesicular membrane is retrieved by endocytosis for reuse. This recycling mechanism¹ ensures a constant supply of releasable vesicles at the nerve terminal¹. The kinetics of endocytosis have been measured mostly after intense or non-physiological stimulation^{2–13}. Here we use capacitance measurements to resolve the fusion and retrieval of single and multiple vesicles following mild physiological stimulation at a mammalian central synapse. The time constant of endocytosis after single vesicle fusion was 56 ms; after a single action potential or trains at ≤ 2 Hz it was about 115 ms, but increased gradually to tens of seconds as the frequency and the number of action potentials increased. These results indicate that an increase in the rate of exocytosis at the active zone induces a decrease in the rate of endocytosis. Existing models^{5,10}, including inhibition of endocytosis by Ca²⁺, could not account for these results—our results suggest that an accumulation of unretrieved vesicles at the plasma membrane slows endocytosis. These findings may resolve the debate about the dependence of endocytosis kinetics on the stimulation frequency^{2,3}, and suggest a potential role of regulation of endocytosis in short-term synaptic depression.

To detect single vesicle endocytosis, both the miniature excitatory postsynaptic current (mEPSC) and the membrane capacitance (C_m) at the calyx of Held (nerve terminal) were simultaneously recorded (Fig. 1a). We increased the mEPSC rate to 24 ± 2 Hz ($n = 46$) by including 10 mM EGTA and 7.5 mM Ca²⁺ in the

presynaptic pipette¹⁴, which clamped the calcium concentration ($[Ca^{2+}]_i$) at 482 ± 29 nM ($n = 3$). As shown later, this manipulation did not affect endocytosis. By averaging 160,776 mEPSCs and the corresponding capacitance traces from 24 synapses, the capacitance jump was detectable (Fig. 1b, left). To improve the signal-to-noise ratio, we averaged 320,541 spontaneous events from 46 synapses (Fig. 1b, right). The capacitance jumped by 65 aF at 1 ms before the onset of the mean mEPSC. Our time resolution was 1 ms because the sinusoidal frequency for capacitance measurements was 1,000 Hz. As the specific membrane capacitance¹⁵ is $9 \text{ fF } \mu\text{m}^{-2}$, 65 aF corresponds to a vesicle with a diameter of 48 nm, which is similar to the estimate from electron microscopy¹⁶. The jump decayed with a time constant (τ) of 56 ms without accompanying parallel changes in the membrane conductance (G_m) and the series conductance (G_s) (Fig. 1b). The decay reflected the mean time course of single vesicle endocytosis.

To investigate endocytosis following action potential stimulation, we searched and found an action potential-equivalent stimulus (AP-e) at the voltage-clamp mode at which the capacitance can be measured. This stimulus was a 1-ms step from -80 to 7 mV, which released a similar number of vesicles as an action potential (Fig. 2a, b; see Methods). The capacitance jump (ΔC_m) evoked by an AP-e or ten AP-e at 333 Hz was mostly blocked ($86 \pm 4\%$, $n = 7$, Fig. 2b and c, or $87 \pm 3\%$, $n = 3$, respectively) by application of the Ca^{2+} channel blocker cadmium ($200 \mu\text{M}$)¹². The jump was not accompanied by parallel changes in G_m or G_s (Fig. 2c, d, f)¹². The decay of the jump following an AP-e was fitted approximately with a single exponential function with $\tau = 115 \pm 3$ ms ($n = 12$, Fig. 2c). Thus endocytosis following single vesicle fusion or single action potentials was 1–2 orders of magnitude faster than the fastest endocytosis estimated following action potential trains (4–20 s)^{8,10,17} or prolonged step depolarization at nerve terminals (0.4–2 s)^{5,7}. Fusion pore flickering, as observed in endocrine cells^{18,19}, could account for such a rapid endocytosis.

To determine whether endocytosis depends on the stimulation frequency, paired AP-e with intervals of 3–2,000 ms were applied. Following the second stimulus at intervals of 500 ms or longer, the endocytic τ was 120 ± 7 ms ($n = 7$, Fig. 2d), similar to that following a single stimulus. However, it gradually increased to 346 ± 19 ms ($n = 13$) as the interval was reduced to 3 ms (Fig. 2d, e), suggesting a frequency-dependent inhibition of endocytosis. In addition, the second capacitance jump was $86 \pm 4\%$ ($n = 13$) and $81 \pm 12\%$ ($n = 13$) of the first one at 20 and 333 Hz, respectively. These decreases appeared less than depression of the EPSC observed at this synapse^{20,21}. Desensitization and/or saturation of postsynaptic AMPA receptors may account for this discrepancy^{12,22}.

The frequency-dependent inhibition of endocytosis applied beyond paired-pulse stimulation. For example, the time constants were 121 ± 8 ms ($n = 6$), 927 ± 26 ms ($n = 6$) and 2.6 ± 0.4 s ($n = 6$) after four AP-e at 2, 20 and 333 Hz, respectively, and

122 ± 9 ms ($n = 6$), 2.3 ± 0.2 s ($n = 8$) and 8.3 ± 0.7 s ($n = 8$) after 10 stimuli at 2, 20 and 333 Hz, respectively (Fig. 2f, g). The time constant increased to 17.3 ± 0.9 s after 20 stimuli at 333 Hz (Fig. 2g). Consistent with previous studies^{5,8,10}, the time constant also depends on the number of stimuli (Fig. 2g).

We now consider how an increase in the frequency and the number of stimuli could slow endocytosis. An increase in the $[Ca^{2+}]_i$ may inhibit endocytosis⁵, so we examined this possibility by using two different pipettes. The first contained the standard solution, with which the basal $[Ca^{2+}]_i$ was 242 ± 28 nM ($n = 4$) and the peak $[Ca^{2+}]_i$ after ten AP-e at 333 Hz was 759 ± 35 nM ($n = 4$, Fig. 3a). As a control, we measured ΔC_m evoked by four and ten AP-e at 333 Hz (Fig. 3c). Then we slowly withdrew the electrode and replaced it with another one containing additional 10 mM EGTA and 8.5 mM calcium^{5,14}. The basal $[Ca^{2+}]_i$ increased to $1.1 \pm 0.2 \mu\text{M}$ ($n = 4$, Fig. 3b), which was higher than the peak $[Ca^{2+}]_i$ induced by ten AP-e in control (Fig. 3a). The ΔC_m induced by ten AP-e was 40–70% smaller than control, and endocytosis was

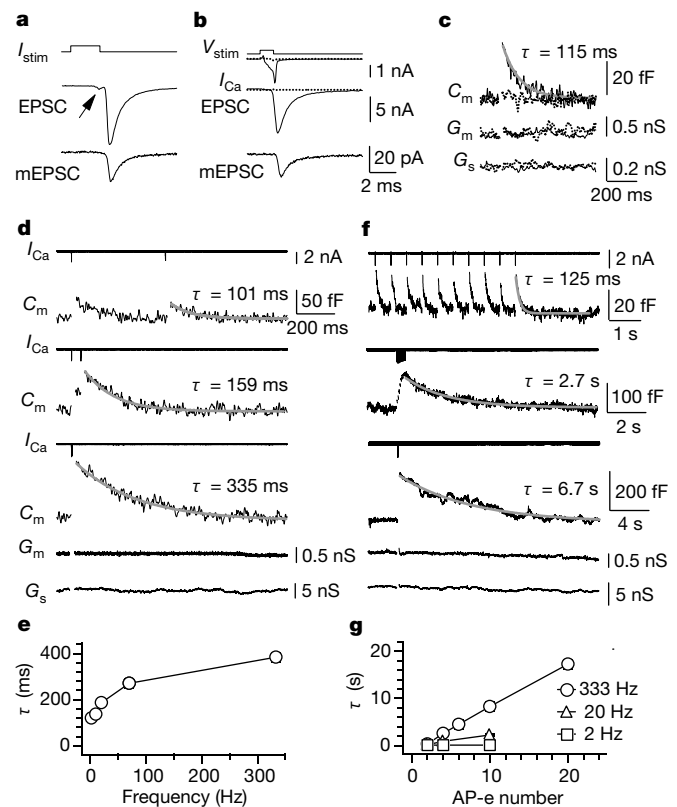


Figure 2 Frequency-dependent inhibition of endocytosis. **a**, In the presynaptic cell-attached configuration, a current injection (I_{stim} , 3 ms, 400 pA) induced an action potential (prespike, arrow) and thus an EPSC. For calculation of the evoked vesicle number, a mean mEPSC was obtained from 47 mEPSCs. **b**, In the presynaptic whole-cell configuration, an AP-e (V_{stim} : 1-ms step from -80 to 7 mV), induced a presynaptic Ca^{2+} current (I_{Ca}) and an EPSC. A mean mEPSC was obtained from 58 mEPSCs. **c**, The C_m , G_m and G_s induced by the AP-e shown in **b**. The decay of C_m was fitted with a single exponential function (grey) with $\tau = 115$ ms. Application of cadmium ($200 \mu\text{M}$) blocked ΔC_m (dotted), I_{Ca} and the EPSC (**b**, dotted). Data in **a–c** were from the same synapse. Capacitance traces were low-pass-filtered at 100 Hz (applies to Figs 2–4). **d**, Sampled I_{Ca} and C_m induced by paired AP-e at 2 (upper), 20 (middle) and 333 (lower) Hz. Grey curves are single exponential fits with τ labelled (applies to **f**). G_m and G_s (lower) were shown only for 333 Hz (applies to **f**). **e**, The endocytic τ after paired AP-e plotted versus the frequency. **f**, Sampled I_{Ca} and C_m (labels as in **d**) induced by ten AP-e at 2 (upper), 20 (middle) and 333 (lower) Hz. The scale for I_{Ca} applies to other traces. **g**, The endocytic τ after a train of AP-e at 2 (squares), 20 (triangles) and 333 (circles) Hz plotted versus the number of AP-e.

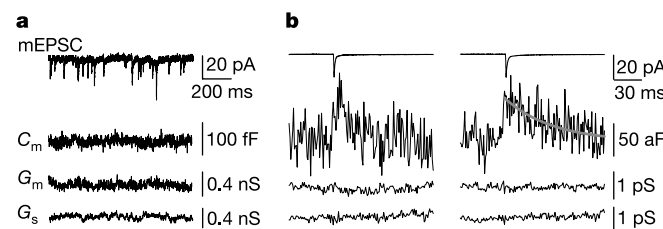


Figure 1 Single vesicle fusion and endocytosis. **a**, Sampled mEPSCs (miniature excitatory postsynaptic currents), presynaptic C_m (membrane capacitance), G_m (membrane conductance) and G_s (series conductance) from a synapse. Labels also apply to **b**. **b**, The mean mEPSC, C_m , G_m and G_s averaged from 160,776 (left) and 320,541 (right, data in left included) spontaneous events. The decay of the capacitance (right) was fitted with a single exponential function (grey) with $\tau = 56$ ms. All traces are shown without filtering.

faster than control (not shown). The smaller ΔC_m was probably caused by EGTA²³ that buffered Ca^{2+} , and an increased spontaneous release that might partly deplete the releasable pool (Fig. 1a)²⁴. We thus increased depolarization in the AP-e by 13–20 mV, which induced a similar ΔC_m as in control. Following four and ten such adjusted AP-e at 333 Hz, the time constants of endocytosis (2.7 ± 0.6 and 7.6 ± 0.5 s, respectively, $n = 5$, Fig. 3d) were similar to control (Figs 2g, 3c), whereas the residual Ca^{2+} transient was essentially abolished by EGTA (Fig. 3b). Similar results ($n = 4$) were obtained with electrodes containing 10 mM EGTA and 9.4 mM Ca^{2+} , which increased the $[Ca^{2+}]_i$ to $2.3 \mu M$ as calculated by assuming one-to-one binding between Ca^{2+} and EGTA with dissociation constant $K_d = 150$ nM (ref. 14). These results suggest that residual Ca^{2+} does not inhibit endocytosis. This conclusion was further confirmed by the result that buffering the residual Ca^{2+} transient by dialysing 10 mM EGTA alone did not significantly change the time constant of endocytosis induced by ten adjusted AP-e at 333 Hz ($n = 4$, not shown).

Is endocytosis inhibited by accumulation of Ca^{2+} in local domains that trigger release? This is unlikely, because local Ca^{2+} domains do not accumulate significantly during a 20-Hz train²⁵, yet the rate of endocytosis depended on the number of stimuli at 20 Hz (Fig. 2g). Does an increased duration of local Ca^{2+} domains inhibit endocytosis? This is also unlikely, because similar time constants of endocytosis (19.4 ± 1.9 s, $n = 7$ and 20.1 ± 3.2 s, $n = 4$, Fig. 3e) were observed following 10- and 30-ms steps to 10 mV, which induced similar capacitance jumps but different durations of Ca^{2+} influx¹². This lack of change was not due to maximal inhibition of endocytosis, because the time constant (41.9 ± 6.8 s, $n = 3$, Fig. 3e) was further increased following five steps of 10 ms to 10 mV at 20 Hz, which increased the capacitance by 907 ± 38 fF ($n = 3$). Thus, unlike the ribbon-type synapse^{5,13}, our results suggest that a mechanism downstream of Ca^{2+} influx regulates endocytosis.

On the basis of Fig. 2g, the endocytic time constant was re-plotted versus the sum of each ΔC_m (the difference between the capacitance value 5–10 ms before and after an AP-e) induced by each AP-e during a train of 1–20 AP-e at 2–333 Hz (Fig. 4a). At 20 and 333 Hz, but not at 2 Hz, the endocytic time constant increased as the summed ΔC_m increased (Fig. 4a). With the same summed ΔC_m , the endocytic time constant increased as the frequency increased (Fig. 4a). These results indicate that endocytosis depends on both

the amount and the rate of exocytosis.

We now consider if there is an intermediate factor by which exocytosis regulates endocytosis. A likely candidate is the net ΔC_m , measured as the difference between capacitance values before and immediately after a train of stimulation. For example, with the same number of AP-e, the net ΔC_m but not the summed ΔC_m increased as the frequency increased from 2 to 333 Hz (Fig. 2d, f). This was because a fraction of fused vesicles had been retrieved during lower frequency trains. The increase in the net ΔC_m paralleled the increase in the endocytic time constant (see, for example, Fig. 2d, f). A strong dependence of the endocytic time constant on the net ΔC_m was revealed when the data obtained during single vesicle fusion, trains of AP-e and prolonged steps, as described in Figs 1–3, were summarized (Fig. 4b). The net ΔC_m was converted to the mean number of remaining fused vesicles per active zone (Fig. 4b) on the basis of the result that a calyx has about 600 active zones¹⁶ or release sites (ranging from 130 to 1,995)²⁶, and that single vesicle fusion increased the capacitance by 65 aF (Fig. 1). A linear regression line with a slope of 1.7 s per vesicle per active zone fits most of the data except the initial part, in which less than one vesicle on average is released per active zone (Fig. 4b). We suggest that the number of unretrieved vesicles, as determined by the amount and the rate of exocytosis, regulates their own rate of retrieval by an as-yet unknown feedback mechanism.

It has been proposed that the endocytic machinery operates at a fixed rate but has a limited capacity¹⁰. Saturation of the machinery may account for slower endocytosis following multiple vesicle fusion¹⁰. This hypothesis predicts (1) the same time course of endocytosis following the same amount of release, (2) a linear capacitance decay during endocytosis, and (3) the same initial velocity of retrieval (number of vesicles per second) following multiple or single vesicle fusion. Our results argue against this hypothesis, because, first, the time course of endocytosis following the same amount of release was different at different stimulation frequencies (Fig. 4a), second, the decay of the capacitance during endocytosis was exponential (Figs 2, 3), and third, the initial velocity of retrieval (0.6–0.7 vesicle per active zone per s) following fusion of 12 vesicles per active zone—calculated as 12 vesicles divided by the endocytic τ (17–20 s, Fig. 4b)—was much slower than that ($1/0.056 = 17.9$ vesicle per active zone per s) following single vesicle fusion (Fig. 4b).

The present work provides several basic parameters of endocytosis at fast synapses. These parameters include the time constant

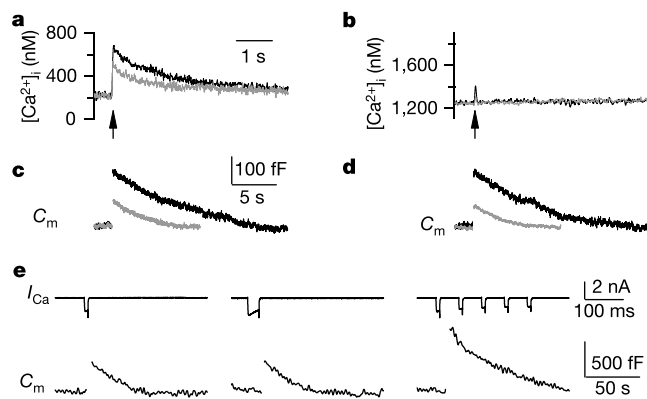


Figure 3 Endocytosis is Ca^{2+} -independent. **a, b**, Sampled Ca^{2+} transients (low-pass-filtered at 30 Hz) induced by four (grey) and ten (dark) AP-e (**a**) or adjusted AP-e (**b**) stimuli (arrow) at 333 Hz in terminals dialysed without (**a**) or with (**b**) 10 mM EGTA and 8.5 mM Ca^{2+} . The timescale bar in **a** applies to **b, c, d**. Sampled C_m induced by four (grey) and ten (dark) AP-e (**c**) or adjusted AP-e (**d**) at 333 Hz with the same pipette solution as in **a** (**c**) or **b** (**d**). Data in **c** and **d** were from the same calyx. Scale bars in **c** apply to **d, e**. Sampled I_{Ca} and C_m induced by a step of 10 (left) or 30 (middle) ms to 10 mV, or by five steps of 10 ms to 10 mV at 20 Hz (right).

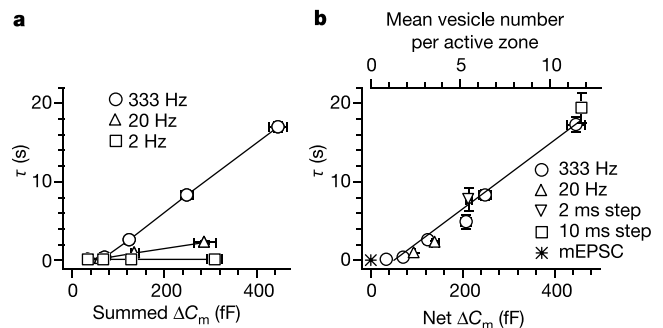


Figure 4 Endocytosis depends on exocytosis. **a**, Endocytic τ plotted versus the sum of the ΔC_m induced by each AP-e during trains at 333 Hz (circles, 1, 2, 4, 10 and 20 AP-e), 20 Hz (triangles, 1, 2, 4 and 10 AP-e) and 2 Hz (squares, 1, 2, 4 and 10 AP-e). **b**, The endocytic τ plotted versus the net ΔC_m (bottom) after stimulation. The stimulation includes trains of AP-e at 333 Hz (circles) and 20 Hz (triangles), 2 ms (inverse triangle) and 10 ms (square) step depolarization, and spontaneous single vesicle fusion (asterisk, mEPSC). The net ΔC_m is also converted to the mean vesicle number per active zone (top, see text). The data were fit with a linear regression line with a slope of 1.7 s per vesicle per active zone.

(56–115 ms) of endocytosis following single vesicle fusion, and single and low frequency (≤ 2 Hz) action potential stimulation. Furthermore, we demonstrate a frequency-dependent but Ca^{2+} -independent inhibition of endocytosis even during paired-pulse stimulation. The amount of unretrieved vesicles, as determined by the balance between the rate of exocytosis and endocytosis, may mediate such a frequency-dependence inhibition. It is counter-intuitive that endocytosis is slower during more intense nerve activity that appears to need rapid vesicle recycling⁹. However, slower endocytosis and thus slower recycling may serve as an inhibitory feedback to prevent excessive vesicle fusion and overstimulation of postsynaptic neurons. Our results call for consideration of potential roles of endocytosis in short-term synaptic plasticity, which lasts for tens to hundreds of milliseconds²⁷. For example, if endocytosed vesicles are preferentially reused¹⁷, reuse will be limited by slow endocytosis during high frequency stimulation, resulting in short-term depression. Consistent with this possibility, short-term depression has been observed during high frequency stimulation at many synapses, including calyceal synapses²⁷.

Three decades ago, Heuser and Reese observed a slow clathrin-dependent endocytosis during 10 Hz nerve stimulation², whereas Ceccarelli and his colleagues proposed a rapid 'kiss and run' endocytosis at active zones on the basis of experiments performed during 2-Hz nerve stimulation³. Today, although both slow and fast endocytosis have been observed^{4–11}, the debate continues on whether the kinetics depends on clathrin⁶ and the stimulation frequency^{9,10}. Our finding of frequency-dependent inhibition of endocytosis provides an explanation for the kinetic discrepancies in the Heuser–Ceccarelli debate. It would be of great interest to understand whether clathrin is involved in mediating different rates of endocytosis. □

Methods

Slice preparation and electrophysiology

Methods for preparing brainstem slices from 8–10-d-old Wistar rats, selection of calyces, and measurements of presynaptic Ca^{2+} currents, capacitance, and AMPA receptor-mediated EPSCs in the medial nucleus of the trapezoid body are described in ref. 12. The $[\text{Ca}^{2+}]_i$ was measured using Fura-2 as described in ref. 28, but with an Olympus upright epifluorescence microscope (BX50WI, LumplanFI 40X, n.a. 0.8) and a polychromatic illumination system (T.I.L.L. Photonics). If not mentioned, recordings were made in a solution that pharmacologically isolated Ca^{2+} currents²⁰. This solution contained (in mM): 105 NaCl, 20 TEA-Cl, 2.5 KCl, 1 MgCl_2 , 2 CaCl_2 , 25 NaHCO_3 , 1.25 NaH_2PO_4 , 25 dextrose, 0.4 ascorbic acid, 3 *myo*-inositol, 2 sodium pyruvate, 0.001 tetrodotoxin (TTX), 0.1 3,4-diaminopyridine, 0.05 D-APV, pH 7.4 when bubbled with 95% O_2 and 5% CO_2 . Bicuculline (10 μM) and strychnine (10 μM) were included when mEPSCs were recorded. The standard presynaptic pipette (2.5–4.5 M Ω) solution contained (in mM): 125 Cs-gluconate, 20 CsCl, 4 MgATP, 10 Na_2 -phosphocreatine, 0.3 GTP, 10 HEPES, 0.05 BAPTA, pH 7.2, adjusted with CsOH. BAPTA was replaced by Fura-2 (50 μM) during recordings of the $[\text{Ca}^{2+}]_i$. When BAPTA was replaced by EGTA and CaCl_2 to increase the $[\text{Ca}^{2+}]_i$, Cs-gluconate concentration was reduced to maintain the osmolarity. Postsynaptic pipette (2–3 M Ω) solution was similar to the presynaptic standard pipette solution with Cs^+ replaced by K^+ and BAPTA replaced by 0.5 mM EGTA. The holding potential for both presynaptic and postsynaptic recordings was -80 mV. For measurements of the capacitance with the EPC-9 amplifier and the software lock-in amplifier (PULSE, HEKA), the frequency of the sinusoidal stimulus added to the holding voltage was 1,000 Hz and the peak-to-peak voltage of the sine wave was less than 60 mV.

Data analysis

Data were expressed as mean $\pm$ s.e.m. When endocytosis lasted for >10 s, the capacitance was recorded for 50 ms every 600 ms in most experiments. Because the presynaptic capacitance was measured during a sinusoidal voltage command at 1,000 Hz, the simultaneously recorded mEPSCs were accompanied by a sinusoidal current (<100 pA) at 1,000 Hz. This noise, detected as the difference between the noises obtained with and without the presynaptic sinusoidal command in the frequency domain, was subtracted. An Igor program (WaveMetrics) was used to detect mEPSCs with amplitudes ≥ 6 –8 pA. The time at the peak of the mEPSC was used to align mEPSCs and corresponding capacitance traces²⁹, and thus to obtain their averaged responses (Fig. 1b).

To search for an AP-e, an electrode was sealed to the calyx in the cell-attached configuration in a bath solution containing no blockers to avoid distortion of the action potential waveform²⁰. This solution contained (in mM): 125 NaCl, 2.5 KCl, 1 MgCl_2 , 2 CaCl_2 , 25 NaHCO_3 , 1.25 NaH_2PO_4 , 25 dextrose, 0.4 ascorbic acid, 3 *myo*-inositol, 2 sodium pyruvate, pH 7.4 when bubbled with 95% O_2 and 5% CO_2 . A 2–4-ms current of 400–1,000 pA was injected to induce a presynaptic action potential and thus an AMPA receptor-mediated EPSC. The EPSC was either 'all or none' depending on whether

the injected current was above a threshold or not. The EPSC was preceded by a prespike (Fig. 2a, arrow) caused via capacitance coupling by the presynaptic action potential²⁰. The EPSC amplitude was 8.5 ± 0.9 nA ($n = 7$), similar to that evoked by single action potentials propagated from the axon²³. These results suggest that the injected current induced single action potentials. Such stimulation released 321 ± 29 ($n = 7$) vesicles, as calculated by dividing the evoked EPSC by the mean mEPSC amplitude (26.5 ± 1.2 pA, $n = 7$, Fig. 2a). Next, we established the presynaptic whole-cell configuration and switched the bath solution to one that pharmacologically isolates Ca^{2+} currents¹². The mEPSC amplitude decreased by $26 \pm 3\%$ ($n = 7$ synapses) in 5 min after break in. Steps of 1 ms to 3–15 mV (7.1 ± 2.0 mV, $n = 7$) released a similar number (318 ± 22 , $n = 7$) of vesicles as that released by an action potential (Fig. 2b). Thus, when postsynaptic recordings were not made, the 1-ms step to 7 mV was used as the AP-e. We did not use an action potential waveform because the resulting EPSC was 2–4 times smaller than that induced by a real action potential²⁸. By scaling the action potential waveform so that the half-width was 0.1–0.3 ms wider, the resulting ΔC_m and decay were similar to that induced by an AP-e ($n = 3$). Furthermore, the Ca^{2+} charge induced by an AP-e was 1.0 ± 0.2 pC ($n = 7$), similar to that induced by an action potential³⁰. These results indicate that an AP-e is equivalent to an action potential.

The number of vesicles released by an AP-e can also be calculated from the evoked ΔC_m (34 fF) and the ΔC_m of single vesicle fusion (65 aF). Taking into account the 14% artefact in the evoked ΔC_m (see text), the vesicle number was 450, which was larger than that (318) estimated using AMPA receptor-mediated EPSCs. This discrepancy might be caused by an overestimate of the mEPSC amplitude, and/or release to postsynaptic sites containing only NMDA receptors.

Measurements of the endocytic time course were not significantly contaminated by spontaneous release, because the number of mEPSCs (frequency <30 Hz) after a stimulus was much less than 10% of evoked release (not shown). The time courses of endocytosis were fitted approximately with single exponential functions (Figs 1–3). Following a stimulus, different active zones may release different numbers of vesicles and thus may retrieve vesicles with different time constants (Fig. 4b). This predicts a multiple exponential function of endocytosis in a calyx containing hundreds of active zones¹⁶. Although some data were fitted better with multiple exponential functions, we did not obtain such data consistently, perhaps because there were too many exponential functions. Thus a single exponential fit may reflect the mean rate of endocytosis at different active zones.

Received 16 January; accepted 4 March 2002.

1. Betz, W. J. & Angleson, J. K. The synaptic vesicle cycle. *Annu. Rev. Physiol.* **60**, 347–363 (1998).
2. Heuser, J. E. & Reese, T. S. Evidence for recycling of synaptic vesicle membrane during transmitter release at the frog neuromuscular junction. *J. Cell Biol.* **57**, 315–344 (1973).
3. Ceccarelli, B., Hurlbut, W. P. & Mauro, A. Turnover of transmitter and synaptic vesicles at the frog neuromuscular junction. *J. Cell Biol.* **57**, 499–524 (1973).
4. Ryan, T. A. *et al.* The kinetics of synaptic vesicle recycling measured at single presynaptic boutons. *Neuron* **11**, 713–724 (1993).
5. Von Gersdorff, H. & Matthews, G. Inhibition of endocytosis by elevated internal calcium in a synaptic terminal. *Nature* **370**, 652–655 (1994).
6. Artalejo, C. R., Henley, J. R., McNiven, M. A. & Palfrey, H. C. Rapid endocytosis coupled to exocytosis in adrenal chromaffin cells involves Ca^{2+} , GTP, and dynamin but not clathrin. *Proc. Natl Acad. Sci. USA* **92**, 8328–8332 (1995).
7. Hsu, S.-F. & Jackson, M. B. Rapid exocytosis and endocytosis in nerve terminals of the rat posterior pituitary. *J. Physiol.* **492**, 539–553 (1996).
8. Wu, L. G. & Betz, W. J. Nerve activity but not intracellular calcium determines the time course of endocytosis at the frog neuromuscular junction. *Neuron* **17**, 769–779 (1996).
9. Klingauf, J., Kavalali, E. T. & Tsien, R. W. Kinetics and regulation of fast endocytosis at hippocampal synapses. *Nature* **394**, 581–585 (1998).
10. Sankaranarayanan, S. & Ryan, T. A. Real-time measurements of vesicle-SNARE recycling in synapses of the central nervous system. *Nature Cell Biol.* **2**, 197–204 (2000).
11. Stevens, C. F. & Williams, J. H. "Kiss and run" exocytosis at hippocampal synapses. *Proc. Natl Acad. Sci. USA* **97**, 12828–12833 (2000).
12. Sun, J. Y. & Wu, L. G. Fast kinetics of exocytosis revealed by simultaneous measurements of presynaptic capacitance and postsynaptic currents at a central synapse. *Neuron* **30**, 171–182 (2001).
13. Neves, G., Gomis, A. & Lagnado, L. Calcium influx selects the fast mode of endocytosis in the synaptic terminal of retinal bipolar cells. *Proc. Natl Acad. Sci. USA* **98**, 15282–15287 (2001).
14. Augustine, G. J. & Neher, E. Calcium requirements for secretion in bovine chromaffin cells. *J. Physiol.* **450**, 247–271 (1992).
15. Gentet, L. J., Stuart, G. J. & Clements, J. D. Direct measurement of specific membrane capacitance in neurons. *Biophys. J.* **79**, 314–320 (2000).
16. Sätzler, K. *et al.* 3-D reconstruction and detailed analysis of active zones located at the calyx of Held. *Soc. Neurosci. Abstr.* **27**, 384.15 (2001).
17. Pyle, J. L., Kavalali, E. T., Piedras-Renteria, E. S. & Tsien, R. W. Rapid reuse of readily releasable pool vesicles at hippocampal synapses. *Neuron* **28**, 221–231 (2000).
18. Alvarez, d. T., Fernandez-Chacon, R. & Fernandez, J. M. Release of secretory products during transient vesicle fusion. *Nature* **363**, 554–558 (1993).
19. Ales, E. *et al.* High calcium concentrations shift the mode of exocytosis to the kiss-and-run mechanism. *Nature Cell Biol.* **1**, 40–44 (1999).
20. Borst, J. G. G., Helmchen, F. & Sakmann, B. Pre- and postsynaptic whole-cell recordings in the medial nucleus of the trapezoid body of the rat. *J. Physiol.* **489**, 825–840 (1995).
21. Von Gersdorff, H., Schneggenburger, R., Weis, S. & Neher, E. Presynaptic depression at a calyx synapse: the small contribution of metabotropic glutamate receptors. *J. Neurosci.* **17**, 8137–8146 (1997).
22. Neher, E. & Sakaba, T. Combining deconvolution and noise analysis for the estimation of transmitter release rates at the calyx of Held. *J. Neurosci.* **21**, 444–461 (2001).
23. Borst, J. G. G. & Sakmann, B. Calcium influx and transmitter release in a fast CNS synapse. *Nature* **383**, 431–434 (1996).
24. Bollmann, J. H., Sakmann, B. & Borst, J. G. Calcium sensitivity of glutamate release in a calyx-type terminal. *Science* **289**, 953–957 (2000).

25. Neher, E. Vesicle pools and Ca^{2+} microdomains: new tools for understanding their roles in neurotransmitter release. *Neuron* **20**, 389–399 (1998).
26. Meyer, A. C., Neher, E. & Schneggenburger, R. Estimation of quantal size and number of functional active zones at the calyx of held synapse by nonstationary EPSC variance analysis. *J. Neurosci.* **21**, 7889–7900 (2001).
27. Von Gersdorff, H. & Borst, J. G. G. Short-term plasticity at the calyx of Held. *Nature Rev. Neurosci.* **3**, 53–64 (2002).
28. Wu, L. G. & Borst, J. G. G. The reduced release probability of releasable vesicles during recovery from short-term synaptic depression. *Neuron* **23**, 821–832 (1999).
29. Chow, R. H., Klingauf, J., Heinemann, C., Zucker, R. S. & Neher, E. Mechanisms determining the time course of secretion in neuroendocrine cells. *Neuron* **16**, 369–376 (1996).
30. Helmchen, F., Borst, J. G. G. & Sakmann, B. Calcium dynamics associated with a single action potential in a CNS presynaptic terminal. *Biophys. J.* **72**, 1458–1471 (1997).

Acknowledgements

We thank S. Mennerick, J. H. Steinbach and R. Wilkinson for comments on the manuscript. This work was supported by the National Science Foundation and the National Institutes of Health.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.G.W. (e-mail: wul@morphus.wustl.edu).

DMY is a Y-specific DM-domain gene required for male development in the medaka fish

Masaru Matsuda*, Yoshitaka Nagahama*, Ai Shinomiya†, Tadashi Sato†, Chika Matsuda*, Tohru Kobayashi*, Craig E. Morrey*, Naoki Shibata‡, Shuichi Asakawa§, Nobuyoshi Shimizu§, Hiroshi Hori||, Satoshi Hamaguchi† & Mitsuru Sakaizumi†

* Laboratory of Reproductive Biology, National Institute for Basic Biology, Okazaki 444-8585, Japan

† Graduate School of Science and Technology, Niigata University, Ikarashi, Niigata 950-2181, Japan

‡ Department of Biology, Faculty of Science, Shinshu University, Asahi 3-1-1, Matsumoto, Nagano 390-8621, Japan

§ Department of Molecular Biology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan

|| Graduate School of Science, Nagoya University, Chikusa-ku, Nagoya 464-8602, Japan

Although the sex-determining gene *Sry* has been identified in mammals¹, no comparable genes have been found in non-mammalian vertebrates. Here, we used recombinant breakpoint analysis to restrict the sex-determining region in medaka fish (*Oryzias latipes*) to a 530-kilobase (kb) stretch of the Y chromosome. Deletion analysis of the Y chromosome of a congenic XY female further shortened the region to 250 kb. Shotgun sequencing of this region predicted 27 genes. Three of these genes were expressed during sexual differentiation. However, only the DM-related² *PGI7* was Y specific; we thus named it *DMY*. Two naturally occurring mutations establish *DMY*'s critical role in male development. The first heritable mutant—a single insertion in exon 3 and the subsequent truncation of *DMY*—resulted in all XY female offspring. Similarly, the second XY mutant female showed reduced *DMY* expression with a high proportion of XY female offspring. During normal development, *DMY* is expressed only in somatic cells of XY gonads. These findings strongly suggest that the sex-specific *DMY* is required for testicular development and is a prime candidate for the medaka sex-determining gene.

The medaka has two major advantages for genetic research: a large genetic diversity within the species^{3,4} and the existence of several inbred strains⁵. As in mammals, sex determination in medaka is male heterogametic⁶, although the Y chromosome is not cytogenetically distinct⁷. Alteration of phenotypic sex with no reproductive consequences, and recombination over the entire sex chromosome pair^{8–10}, suggest that there are no major differences, other than a sex-determining gene, between the X and Y chromosomes. To clone positionally the sex-determining region, we generated a Y congenic strain to highlight the genetic differences between the X and Y chromosomes from inbred strains of medaka¹¹. The Y congenic strain has a sex-determining region derived from the HNI-strain Y chromosome on the genetic background of an Hd-rR strain. In this congenic strain, the wild-type allele (*R*) of the *r* locus (a sex-linked pigment gene) is located only on the Y chromosome. Therefore, the female X^rX^r results in a white body colour, and the male X^rY^R results in an orange-red body colour. Using this strain, we had previously constructed a genetic map of the medaka sex chromosome⁹.

In this study, we first performed chromosome walking using several recombinants to map the sex-determining region of the Y congenic strain (Fig. 1a). Two recombinants (Hd-rY.^{HNI}(R1) and (R2)) between the sex-determining (*SD*) locus and a sex-linked marker (*SL1*) (centromere side of *SD*) were obtained from an oestrogen-induced XY female of the Y congenic strain⁹. To obtain recombinants between *SD* and *r* (a body-colour gene), we re-screened progeny from an oestrogen-induced XY female of the recombinant Hd-rY.^{HNI}(R1) strain. We subsequently found one white male (Hd-rY.^{HNI}*rr*) and one orange-red female, which were recombinants between the *SD* and *r* loci. After confirming their genotypes from fin clippings, these recombinants were maintained as strains.

For chromosome walking from *SL1*, we constructed a bacterial artificial chromosome (BAC) genomic library from the Y congenic strain. By this approach, we obtained 47 sex-linked BAC clones, sequenced their end fragments, and designed polymerase chain reaction (PCR) primers for sequence tag site (STS) markers on the Y chromosome. Analyses of recombinant genotypes with these STS markers indicated that the sex-determining region was located between *135D12.F* and *51H7.F* (Fig. 1a). This stretch of the Y chromosome was encompassed by four BAC clones (Fig. 1b), although 19 other BAC clones included some portion of this region.

We then determined the location of the sex-determining region on the Y chromosome by fluorescence *in situ* hybridization (FISH) using one of the BAC clones as a probe. The sex-determining region is located on the centromere side of the long arms of the sex chromosomes (Fig. 1c, d). Centromere and DNA marker locations on the sex chromosomes, determined using triploid hybrid¹² and gynogenetic diploids¹⁰, confirmed the findings of the FISH analysis.

We used shotgun sequencing to determine the sequence covered by the four BAC clones. The entire sequence of the two centromere-side BAC clones (mCON089P3 and mCON144M14) was determined; however, the remaining two, telomere-side, BAC clones (mCON104P2 and mCON137M1) could not be completely sequenced mainly owing to numerous repetitive sequences. Consequently, we sequenced 422,202 nucleotides and estimated that the four BAC clones covered about 530 kb (Fig. 1b). The gene-predicting program Genscan (Version 1.0, <http://genes.mit.edu/GENSCAN.html>) predicted 52 genes in this region.

We also found an orange-red female in our congenic progeny (Fig. 2). Mating this female with a sex-reversed (androgen-induced) XX Hd-rR male resulted in all female offspring with a 50:50 (white:orange-red) body colour ratio. Assuming this female's X chromosome was derived from a recombination event between the *SD* and *r* loci, sex-linked DNA markers flanking the *SD* locus should be homozygous (Hd-rR/Hd-rR type). The markers (for example, *SL1* and *51H7.F*) flanking the *SD* were heterozygous (Hd-rR/HNI

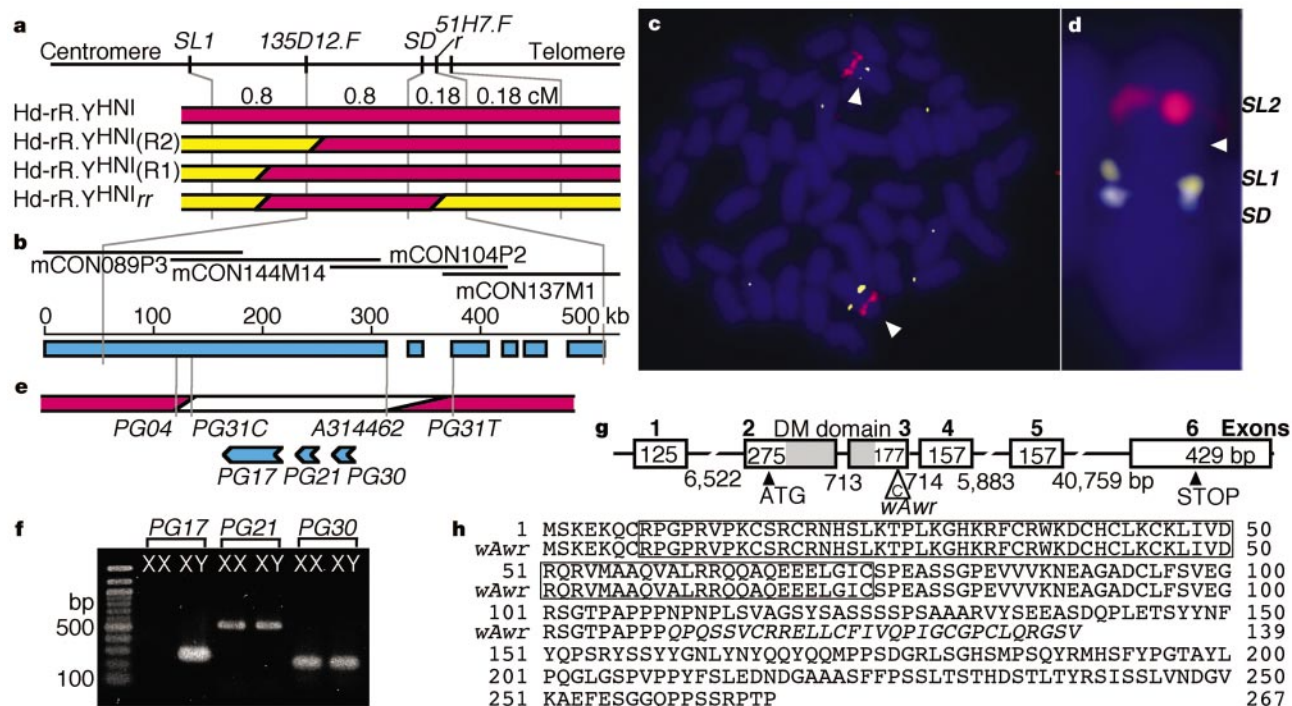


Figure 1 Positional cloning strategy of the sex-determining region and subsequent identification of PG17/DMY. **a**, Genetic maps of the sex-determining regions and Y chromosomes of the recombinant strains. Pink indicates regions derived from an HNI Y chromosome; yellow represents Hd-rR X-chromosome-derived regions. **b**, BAC contigs and a physical map of the sex-determining region. Horizontal bars indicate BAC clones. Blue blocks show the sequenced regions. **c**, **d**, Cytogenetical mapping of the sex-determining region of medaka. **c**, FISH of metaphase chromosomes. SL2 (red) localizes on the short arms of the sex chromosomes, whereas a BAC clone containing SL1 (mCON072N5, yellow) localizes on the long arms. Arrowheads indicate sex chromosomes. **d**, FISH of one sex chromosome with three different probes (SL2, SL1, SD). Signals of a BAC clone containing the sex-determining region (mCON049E13) are light blue. The arrowhead shows the centromere region. **e**, The Y chromosome of medaka lacking a part of the sex-determining region. DNA markers were not detected between

PG17 and PG31T, but were detected on the centromere side of PG04 and the telomere side of PG31T. Respective location of PG17, PG21 and PG30 within the sex-determining region between PG04 and PG31T is shown. **f**, RT-PCR analysis for PG17, PG21 and PG30 mRNA expression in whole Hd-rR.Y^{HNI} embryos at hatching. Sex chromosome types were determined by SL1 PCR. PG17 expression was detected by nested PCR (see Methods). **g**, **h**, Structure and sequences of PG17/DMY. **g**, Structural analysis of the PG17/DMY showing exons (open boxes), the DM domain (grey boxes) and introns (horizontal lines) of PG17. Translation start and stop sites are indicated by ATG and STOP, respectively. Numbers represent nucleotide sequence length (base pairs, bp). **h**, Amino-acid sequence of wild-type PG17/DMT and the PG17^{wawr} mutation. The DM domain is boxed. A frame shift (italicized) occurs in the region from amino acid 110 to premature termination (amino acid 139).

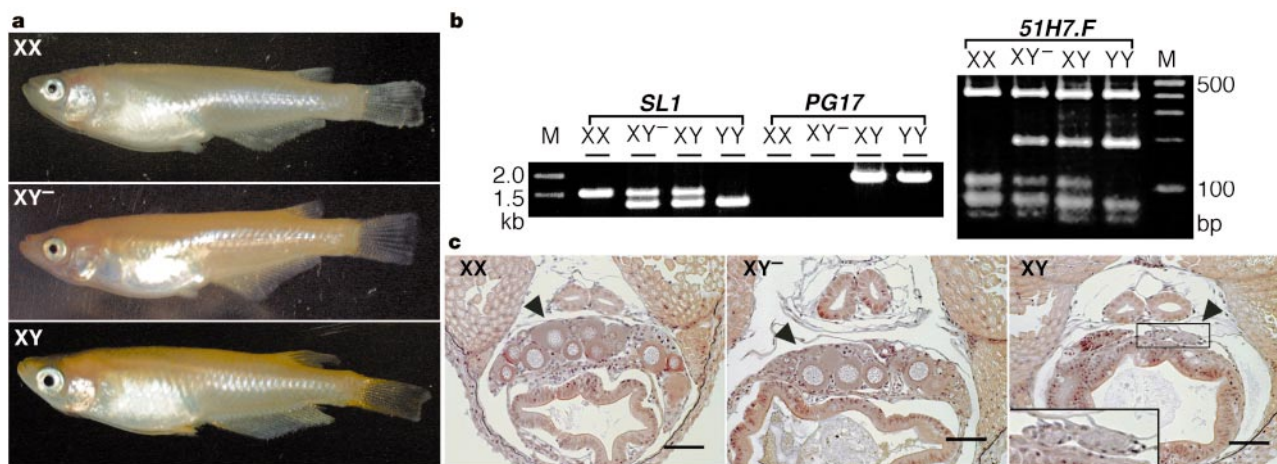


Figure 2 Characteristics of medaka lacking a part of the Y chromosome. **a**, Phenotypes of the congenic strain (XX, XY) and medaka lacking a part of the Y chromosome (XY⁻). XX (top), XY⁻ and XY (bottom) with their secondary sex characters. Males (XY) have larger anal and dorsal fins than females (XX, XY⁻). XX have white bodies (X^X), whereas the others are orange-red (X^Y). **b**, DNA types (SL1, PG17 and 51H7.F) of sex chromosomes. PCR products of SL1 and PG17 were electrophoresed in a 1% agarose gel with a 1-kb DNA ladder. *A*lu-digested PCR products of 51H7.F were electrophoresed in a 3% agarose gel with a 100-bp DNA ladder. SL1 and 51H7.F are homozygous in XX and YY (Hd-rR or HNI type), but are heterozygous in XY⁻ and XY (Hd-rR/HNI type). PG17 is present in XY and

YY but absent in XX and XY⁻. These results indicate that PG17 is specific to the Y chromosome but absent from the Y chromosome of XY⁻. Specific primers for PG17 and 51H7.F were as follows: PG17: PG17.19, GAACACAGCTTGAAGACCCCGCTGA; PG17.20, GCATCTGCTGGTACTGCTGGTAGTTG; and 51H7.F: 51H7.F2, CAGGCCCTGAAGATCAACGAGT; 51H7.F3, AGTGCATCTAGTGACATGGGT. **c**, Histological cross-sections of medaka fry sampled 30 d.a.h.. Sex chromosome types were determined by PCR using DNA extracted from the head. Black arrowheads indicate gonads. XX and XY⁻ individuals have ovaries with several oocytes, whereas XY specimens have testes with spermatogonia. Scale bars, 50 μ m.

Table 1 Genotype and gonadal phenotype ratios

Cross	Genotype	Gonadal phenotype ratio (ovary:testis)	
		Fry	Adults
XY ^{wAw} (female) × Hd-rR (male)	XX	7: 0	11: 0
	XY ^{wAw}	10: 0	8: 0
	XY ^{Hd-rR}	0: 11	0: 5
	YY	0: 8	0: 2
XY ^{wSm} (female) × Hd-rR (male)	XX	NA	15: 0
	XY ^{wSm}	NA	9: 6
	XY ^{Hd-rR}	NA	0: 9
	YY	NA	0: 8

Ratios are from progeny of mutant female × Hd-rR male crosses. Sex chromosome type was determined from *SL1* and *PG17* genotypes. NA, not applicable (fry were not analysed).

type), indicating that this particular orange-red female had the congenic sex-determining region on the Y chromosome (Fig. 2b), a conclusion confirmed by other DNA markers within the sex-determining region. From these observations, this XY female was determined to lack 250 kb within the sex-determining region of the Y chromosome (Fig. 1e). Consequently, the corresponding region of a normal Y chromosome should contain the sex-determining gene.

Genscan analysis of the deleted 250-kb region predicted 27 genes. To identify which predicted genes were expressed, we designed specific primers for each, and examined their expression in medaka embryos during sexual differentiation (before hatching to 10 days after hatching, d.a.h.). PCR with reverse transcription (RT-PCR) detected expression of three of these 27 genes in embryos (Fig. 1f). Furthermore, only one of these three genes, *PG17* (predicted gene 17)/*DMY*, was expressed exclusively in XY embryos. In contrast, the remaining two genes (*PG21*, *PG30*) were expressed in both XY and XX embryos. BAC clones derived from the X and Y chromosomes confirmed that *PG17* is specific to the Y chromosome whereas the other two genes are present on both the X and Y chromosomes.

The full-length complementary DNA sequence (1,320 base pairs) of *DMY* was obtained by 5' and 3' rapid amplification of cloned ends (RACE). The longest open reading frame spans six exons and encodes a putative protein of 267 amino acids, including the highly

conserved DM domain (Fig. 1g, h). The DM domain was originally described as a DNA-binding motif shared between *doublesex* (*dsx*) in *Drosophila melanogaster* and *mab-3* in *Caenorhabditis elegans*². Since the initial characterizations of *dsx* and *mab-3*, DM-related genes have been identified from virtually all species examined, including medaka^{13–19}. In vertebrate species, *DMRT1* (DM-related transcription factor 1), the DM-related gene most homologous to *DMY* (about 80%; data not shown), correlates with male development^{13–18,20}. Combined with its Y chromosome specificity, this finding suggests that *DMY* is pivotal in testicular differentiation.

To establish a role for *DMY* during sexual differentiation, we screened wild medaka populations for naturally occurring *DMY* mutants. Two XY females with distinct mutations in *DMY* were found in separate populations (Awara and Shirone). The XY^{wAw} mutant from Awara contains a single nucleotide insertion in exon 3 of *DMY*. Although the DM domain remains intact, this insertion causes a frame shift from residue 110 and premature termination at residue 139 (Fig. 1h). Offspring obtained by mating the XY^{wAw} female with an Hd-rR male (XY, *DMY* on Y chromosome) revealed typical mendelian combinations (XX, XY, XY^{wAw}, YY^{wAw}); however, phenotypes were female, male, female and male, respectively (Table 1). Despite the presence of pseudo-*DMY* messenger RNA (which does not encode full-length *DMY*) in offspring containing the *PG17*^{wAw} mutant allele (Fig. 1h), the absence of about two-thirds of the protein presumably renders *DMY* nonfunctional, thus resulting in XY sex reversal (female phenotype). Although the mechanism apparently differs, the second mutant also results in XY sex reversal. The entire *DMY* coding region of the XY^{wSm} mutant is intact; however, an unknown transcriptional anomaly severely depresses or eliminates *DMY* expression in embryos (Fig. 3). In addition to the original XY^{wSm} female mutant, 60% of XY^{wSm} offspring developed as phenotypic females (Table 1), suggesting that a threshold level of *DMY* expression is required for male development. Taken together, the loss-of-function mutant (XY^{wAw}) and the depressed expression mutant (XY^{wSm}) strongly suggest that *DMY* is

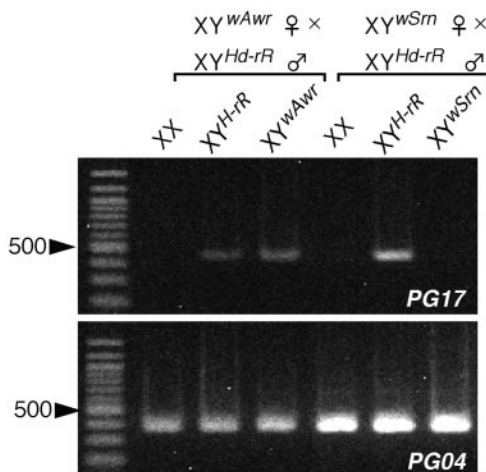


Figure 3 cDNA PCR of progeny from XY^{wAw} and XY^{wSm} female × Hd-rR male crosses. A *PG17/DMY* band was detected in XY^{Hd-rR} and XY^{wAw} embryos but not in XY^{wSm} embryos. Total RNA (150 ng each) was extracted from whole embryos (at hatching) and used for cDNA synthesis and amplification using the SMART PCR cDNA Synthesis Kit (Clontech) according to the manufacturer's protocol. Amplified cDNA and genomic DNA were used as PCR templates of *PG04* and *PG17* (19–20 primer set). Specific primers for *PG04* were as follows: *PG04.1*, CCAGCGTTTGAGGATAGGTTTG; *PG04.2*, GAGCTTTCTGCAGGGCGACTTTC. Products were electrophoresed in a 2% agarose gel with a 100-bp DNA ladder.

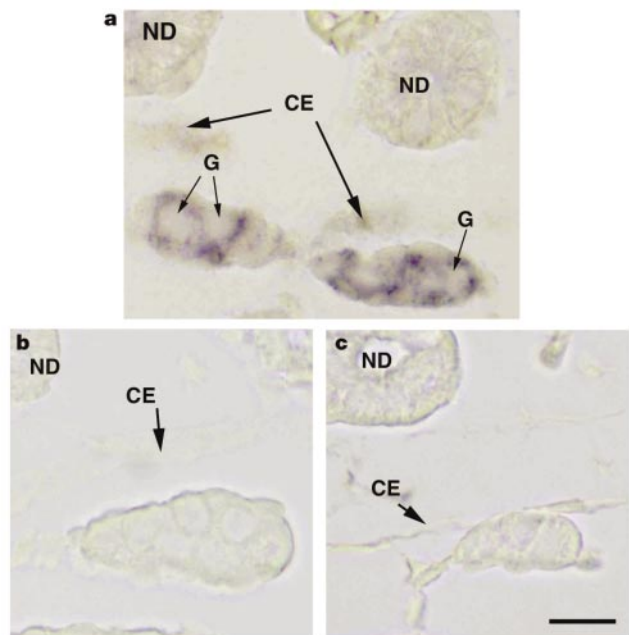


Figure 4 *PG17/DMY* mRNA expression in fry gonads shown by digoxigenin-labelled *in situ* hybridization of larval sections. **a**, XY gonads on hatching day; antisense probe. Strong signals for *PG17/DMY* were seen in cells surrounding the germ cells. G, germ cell; CE, coelomic epithelium; ND, nephric duct. **b**, XX gonad on hatching day; antisense probe. *PG17/DMY* signal was undetectable in XX individuals. **c**, XY gonad on hatching day; sense probe. Control hybridization showed no signal. Scale bar, 20 μ m.

required for normal testicular differentiation.

To confirm the role of *DMY* during normal development, *in situ* hybridization was used to determine its spatial expression during gonadal differentiation. Because *DMRT1* and *DMY* are very similar, the *in situ* hybridization probe was not able to discriminate between the two mRNAs. We therefore determined the temporal expression of both *DMY* and *DMRT1* during sexual differentiation using specific RT-PCR. At hatching and 5 d.a.h., *DMY* mRNA was present in XY embryos, but not in XX embryos. In contrast, there was no *DMRT1* expression in either XX or XY embryos at either of these time points (data not shown). Because expression of *DMY* and *DMRT1* do not appear to overlap during this period, we assumed that the *in situ* hybridization signal represented *DMY* expression. As predicted from our PCR studies, *DMY* signal was detected only in the somatic cells surrounding germ cells in XY embryos (Fig. 4). Although its function in the pre-Sertoli cells remains unclear, these data further indicate that *DMY* has a critical role in testicular differentiation.

Although evidence of sufficiency is required to definitively identify *DMY* as a sex-determining gene, we have shown that it is necessary for normal male development and falls within the sex-determining region of the Y chromosome. Given the absence of other reasonable candidate genes within the region, *DMY* is certainly the leading candidate for the medaka sex-determining gene. Interestingly, phylogenetic analyses indicate that *DMY* was probably derived from *DMRT1*, suggesting an evolutionary pattern similar to one of the proposed origins of *Sry*. *Sry* is thought to have either arisen from an autosomal *Sox* gene duplication event and subsequent formation of the Y chromosome or, more probably, divergence from *Sox3* after formation of the sex (X) chromosomes²¹. Regardless, the linkage of *Sry* to the Y chromosome renders the entire male pathway dependent on *Sry*. Further evidence concerning the ability of *DMY* to trigger male development and its evolutionary relationship to *DMRT1* are needed to confirm this hypothesis, but the linkage of *DMY* to the Y chromosome and its requirement for testicular differentiation strongly suggest that *DMY* represents a non-mammalian vertebrate equivalent of *Sry*. □

Methods

Fish

Two recombinant strains between *SL1* and *SD*, Hd-rY^{HNI} (R1) and Hd-rY^{HNI} (R2), were established from recombinant offspring of sex-reversed XY females of the described Hd-rY^{HNI} strain⁹. Another recombinant strain, Hd-rY^{HNI}rr, was established from a recombinant between *SD* and *r*. This recombinant was obtained from offspring of a sex-reversed Hd-rY^{HNI} (R1) XY female crossed with a sex-reversed Hd-rY^{HNI} XX male. Sex reversal in medaka was accomplished by oestrogen (XY females) or androgen (XX males) treatment during early development, according to previously published methods⁹. Naturally occurring mutants XY^{wAw} and XY^{wSm} were found in wild populations near Awara (Fukui prefecture) and Shirone (Niigata prefecture), Japan, respectively.

RNA and DNA extraction

Total RNA and genomic DNA were extracted from each hatched embryo after homogenization in a 1.5-ml tube with 350 µl RLT buffer supplied with the RNeasy Mini Kit (Qiagen). The homogenized lysates were centrifuged and supernatants were used for RNA extraction using the RNeasy Mini Kit with the RNase-Free DNase set protocol (Qiagen). Precipitated material was used for DNA extraction using the DNeasy tissue Kit (Qiagen) according to the manufacturer's protocol.

Chromosome walking

A BAC genomic library was constructed from the Y congenic Hd-rY^{HNI} strain as described²². High-molecular-mass genomic DNA was extracted from sperm, partially digested with *HindIII*, and selected for a size range of 150–250 kb. The size-selected DNA fragments were ligated to pBAC-lac²³ and used to transform DH10B. A total of 55,292 BAC clones was picked and arrayed to 144 microtitre plates each with 384 wells. The library was gridded at high density on Hybond-N + nylon membranes (Amersham Pharmacia Biotech). Chromosome walking started at *SL1*. Two-thirds of the library was usually screened. Inserted end fragments of positive BAC clones were amplified by vectorette PCR²⁴ and used for assembling the positive clones. An amplified end fragment at the far end of the *SD* side was used in subsequent screening of the BAC library.

RT-PCR

First-strand cDNA was synthesized from 0.5 µg total RNA in 25 µl using PowerScript (Clontech) with oligo-dT primers. PCR was carried out in a 25 µl reaction mixture containing 0.25 µl of the first-strand cDNA. PCR conditions were 5 min at 95 °C; 20 s at

96 °C, 30 s at 55 °C, 30 s at 72 °C for 30 cycles; and 5 min at 72 °C. For *PG17*, 0.01 µl of the initial PCR products were re-amplified under the same conditions. Specific primers for each gene were as follows: *PG17*—first *PG17.11*, GAGTCGGAGCCAAGCGGGTACAA CATTG, *PG17.12*, GACCATCTCATTCTTTTATCTTGATTCT, second *PG17.5*, CCGGGTGCCCAAGTGCTCCGCTG, *PG17.6*, GATCGTCCCTCCACAGAGAAGAGA; *PG21*—*PG21.1*, TGTGATTCTGAAGGGGGAGTTTGTA, *PG21.3*, GACCTCCAGAGTC ATCTTGACACAC; and *PG30*—*PG30.1*, GGAGGAAAGTGTCAGAGTGTGTGT, *PG30.2*, GCCGTCCCTCTGATGTACTCGTTCCT).

FISH mapping

Metaphase cells from Hd-rY embryos were prepared by standard cytogenetic methods²⁵. FISH was performed using an *SL2* fragment labelled by PCR⁹ with digoxigenin-11-dUTP (Boehringer) and BAC clones mCON072N5 (containing *SL1*) and mCON049E13 (*SD* region) labelled by nick translation with fluorescein and biotin, respectively. Cells were co-hybridized with *SL2* and *SL1* or *SL2*, *SL1* and *SD*, and counterstained with 4,6-diamidino-2-phenylindole (DAPI). Probes were detected with rhodamine-labelled anti-digoxigenin antibodies (Boehringer), Alexa Fluor 488-labelled anti-fluorescein, and Alexa Fluor 660-labelled streptavidin (Molecular Probes). Metaphase cells were examined using four filters (A4, L5, N3 and Y5) and images were captured using a CoolSNAP charge-coupled device camera (Nippon roper) and Openlab software (Improvision). Hybridization was detected on the identical sex chromosome.

Shotgun sequencing

BAC DNA was hydrodynamically sheared to average sizes of 1.5 and 4.5 kb, and the DNA was ligated into a pUC18 vector. We sequenced each BAC to 13 coverage using Dye-terminator chemistry. Individual BACs were assembled from the shotgun sequences using phred Version 0.000925.c, crossmatch Version 0.990319 and phrap Version 0.990319 (Codon Code), and PCP Version 2.1.6 and CAP4 Version 2.1.6 (Paracel). The gaps in each BAC were closed using a combination of BAC walking, directed PCR and re-sequencing of individual clones.

Analyses of wild mutants

We crossed wild mutant females (XY; $-/PG17^{wAw}$ or XY; $-/PG17^{wSm}$) with Hd-rY males (XY; $-/PG17^{Hd-rY}$) to obtain the following sex chromosomes and *PG17* genotypes in the offspring: XX; $-/-$, XY; $-/PG17^{wAw}$, XY; $-/PG17^{Hd-rY}$, and YY; $PG17^{wAw}/PG17^{Hd-rY}$. Genotypes of these offspring were determined by genomic PCR of *PG17* and *SL1*. *PG17* expression was examined by RT-PCR (see above). Sexes were confirmed by histological examination of gonads (fry and adult fish) and secondary sex characteristics (adult fish). At hatching, germ cell numbers were counted to determine gonadal phenotype²⁶. Exon sequences of the mutant *PG17* genes (*PG17^{wAw}* and *PG17^{wSm}*) were determined using DNA extracted from the mutants and their progeny.

In situ hybridization

Gonads of 0–5 d.a.h. Hd-rY fry were dissected with trunk of the body and fixed in 4% paraformaldehyde in 0.1 M phosphate buffer (pH 7.4) at 4 °C overnight. After fixation, gonads were embedded in paraffin and cross-sectioned at 5 µm. *In situ* hybridization was performed using published methods²⁷. Genetic sex of fry was confirmed by PCR using the *SL1* genotype¹¹.

Received 22 February; accepted 24 April 2002.

Published online 12 May 2002, DOI 10.1038/nature751.

1. Sinclair, A. H. *et al.* A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346**, 240–244 (1990).
2. Raymond, C. S. *et al.* Evidence for evolutionary conservation of sex-determining genes. *Nature* **391**, 691–695 (1998).
3. Sakaizumi, M., Moriawaki, K. & Egami, N. Allozymic variation and regional differentiation in wild populations of the fish *Oryzias latipes*. *Copeia* **1983**, 311–318 (1983).
4. Matsuda, M., Yonekawa, H., Hamaguchi, S. & Sakaizumi, M. Geographic variation and diversity in the mitochondrial DNA of the medaka, *Oryzias latipes*, as determined by restriction endonuclease analysis. *Zool. Sci.* **14**, 517–526 (1997).
5. Hyodo-Taguchi, Y. & Sakaizumi, M. List of inbred strains of the medaka, *Oryzias latipes*, maintained in the Division of Biology, National Institute of Radiological Sciences. *Fish Biol. J. MEDAKA* **5**, 29–30 (1993).
6. Aida, T. On the inheritance of colour in a freshwater fish, *Aplocheilichthys latipes* Temminck and Schlegel, with special reference to sex-linked inheritance. *Genetics* **6**, 554–573 (1921).
7. Uwa, H. & Ojima, Y. Detailed and banding karyotype analysis of the medaka, *Oryzias latipes*, in cultured cells. *Proc. Jpn Acad. B* **57**, 39–43 (1981).
8. Yamamoto, T. Progenies of sex-reversal females mated with sex-reversal males in the medaka, *Oryzias latipes*. *J. Exp. Zool.* **146**, 163–179 (1961).
9. Matsuda, M., Sotoyama, S., Hamaguchi, S. & Sakaizumi, M. Male-specific restriction of recombination frequency in the sex chromosomes of the medaka, *Oryzias latipes*. *Genet. Res.* **73**, 225–231 (1999).
10. Kondo, M., Nagao, E., Mitani, H. & Shima, A. Differences in recombination frequencies during female and male meioses of the sex chromosomes of the medaka, *Oryzias latipes*. *Genet. Res.* **78**, 23–30 (2001).
11. Matsuda, M. *et al.* Isolation of a sex chromosome-specific DNA sequence in the medaka, *Oryzias latipes*. *Genes Genet. Syst.* **72**, 263–268 (1997).
12. Sato, T., Yokomizo, S., Matsuda, M., Hamaguchi, S. & Sakaizumi, M. Gene-centromere mapping of medaka sex chromosomes using triploid hybrids between *Oryzias latipes* and *O. luzonensis*. *Genetica* **111**, 71–75 (2001).
13. Raymond, C. S., Kettlewell, J. R., Hirsch, B., Bardwell, V. J. & Zarkower, D. Expression of *Dmrt1* in the genital ridge of mouse and chicken embryos suggests a role in vertebrate sexual development. *Dev. Biol.* **215**, 208–220 (1999).
14. Smith, C. A., McClive, P. J., Western, P. S., Reed, K. J. & Sinclair, A. H. Conservation of a sex-

determining gene. *Nature* **402**, 601–602 (1999).

15. De Grandi, A. *et al.* The expression pattern of a mouse doublesex-related gene is consistent with a role in gonadal differentiation. *Mech. Dev.* **90**, 323–326 (2000).
16. Kettlewell, J. R., Raymond, C. S. & Zarkower, D. Temperature-dependent expression of turtle *Dmrt1* prior to sexual differentiation. *Genesis* **26**, 174–178 (2000).
17. Marchand, O. *et al.* *DMRT1* expression during gonadal differentiation and spermatogenesis in the rainbow trout, *Oncorhynchus mykiss*. *Biochim. Biophys. Acta* **1493**, 180–187 (2000).
18. Guan, G., Kobayashi, T. & Nagahama, Y. Sexually dimorphic expression of two types of DM (*Doublesex/Mab-3*)-domain genes in a teleost fish, the tilapia (*Oreochromis niloticus*). *Biochem. Biophys. Res. Commun.* **272**, 662–666 (2000).
19. Brunner, B. *et al.* Genomic organization and expression of the doublesex-related gene cluster in vertebrates and detection of putative regulatory regions for *dmrt1*. *Genomics* **1**, 8–17 (2001).
20. Moniot, B., Berta, P., Scherer, G., Sudbeck, P. & Poulat, F. Male specific expression suggests role of *DMRT1* in human sex determination. *Mech. Dev.* **91**, 323–325 (2000).
21. Schepers, G. & Koopman, P. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Dev. Biol.* **227**, 239–255 (2000).
22. Matsuda, M. *et al.* Construction of a BAC library derived from the inbred Hd-rR strain of the teleost fish, *Oryzias latipes*. *Genes Genet. Syst.* **76**, 61–63 (2001).
23. Asakawa, S. *et al.* Human BAC library: construction and rapid screening. *Gene* **191**, 69–79 (1997).
24. Ragoussis, J. & Olavesen, M. G. in *Genome Mapping: A Practical Approach* (ed. Dear, P. H.) 253–260 (Oxford Univ. Press, New York, 1997).
25. Matsuda, M., Matsuda, C., Hamaguchi, S. & Sakaizumi, M. Identification of the sex chromosomes of the medaka, *Oryzias latipes*, by fluorescence *in situ* hybridization. *Cytogenet. Cell Genet.* **82**, 257–262 (1998).
26. Hamaguchi, S. A light- and electron-microscopic study on the migration of primordial germ cells in the teleost, *Oryzias latipes*. *Cell Tissue Res.* **227**, 139–151 (1982).
27. Kobayashi, T., Kajiura-Kobayashi, H. & Nagahama, Y. Differential expression of *vasa* homologue gene in the germ cells during oogenesis and spermatogenesis in a teleost fish, tilapia, *Oreochromis niloticus*. *Mech. Dev.* **99**, 139–142 (2000).

Acknowledgements

We are grateful to P. Koopman for advice; G. Young for critical reading of the manuscript; and M. Takeda, E. Uno and R. Hayakawa for technical assistance. This work was supported in part by Grants-in-Aid for Research for the Future from the Japan Society for the Promotion of Science, Scientific Research of Priority Area, Environmental Endocrine Disrupter Studies from the Ministry of the Environment, Bio Design from the Ministry of Agriculture, Forestry and Fisheries, and Japan Society for the Promotion of Science Research Fellowships for Young Scientists.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.N. (e-mail: nagahama@nibb.ac.jp). The DNA Data Bank of Japan (DDBJ) accession number of the medaka *DMY* cDNA sequence is AB071534.

Transcription coactivator TRAP220 is required for PPAR γ 2-stimulated adipogenesis

Kai Ge*, Mohamed Guermah*, Chao-Xing Yuan*[‡], Mitsuhiro Ito*, Annika E. Wallberg*, Bruce M. Spiegelman[†] & Robert G. Roeder*

* Laboratory of Biochemistry and Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

[†] Dana-Farber Cancer Institute and Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

The TRAP (thyroid hormone receptor-associated proteins) transcription coactivator complex (also known as Mediator) was first isolated as a group of proteins that facilitate the function of the thyroid hormone receptor¹. This complex interacts physically with several nuclear receptors through the TRAP220 subunit, and with diverse activators through other subunits². TRAP220 has been reported to show ligand-enhanced interaction with peroxisome proliferator-activated receptor γ 2 (PPAR γ 2)^{3,4}, a nuclear receptor essential for adipogenesis^{5–8}. Here we show

that *Trap220*^{−/−} fibroblasts are refractory to PPAR γ 2-stimulated adipogenesis, but not to MyoD-stimulated myogenesis, and do not express adipogenesis markers or PPAR γ 2 target genes. These defects can be restored by expression of exogenous TRAP220. Further indicative of a direct role for TRAP220 in PPAR γ 2 function via the TRAP complex, TRAP functions directly as a transcriptional coactivator for PPAR γ 2 in a purified *in vitro* system and interacts with PPAR γ 2 in a ligand- and TRAP220-dependent manner. These data indicate that TRAP220 acts, via the TRAP complex, as a PPAR γ 2-selective coactivator and, accordingly, that it is specific for one fibroblast differentiation pathway (adipogenesis) relative to another (myogenesis).

PPAR γ is a key regulator of transcriptional pathways important for adipogenesis⁸. Retrovirus vector-mediated ectopic expression of PPAR γ 2 alone can stimulate mouse embryonic fibroblasts (MEFs) to undergo adipogenesis^{9,10}. PPAR γ is a member of the nuclear hormone receptor superfamily of DNA binding transcriptional activators that, in a ligand-dependent manner, activate specific target genes important for cell growth, cell differentiation and homeostasis. Like other transcriptional activators, nuclear receptors act through a variety of interacting transcriptional coactivators that act either to modify chromatin structure or at subsequent steps such as preinitiation complex (PIC) assembly and function^{2,11}. The mammalian TRAP complex, like the distantly related yeast Mediator, appears to play the main role in direct communication between diverse activators and the general transcription factors that form the PIC^{2,12–19}. The demonstration of ligand-dependent interactions of the TRAP220 subunit of the TRAP complex with multiple nuclear receptors suggested a broad role for TRAP in nuclear receptor function and associated physiological processes⁴. As the TRAP220 subunit of the TRAP complex has been shown to interact physically with PPAR γ 2^{3,4}, we sought to investigate the physiological role of TRAP220, and the possible involvement of the entire TRAP complex, in PPAR γ 2-stimulated adipogenesis and target

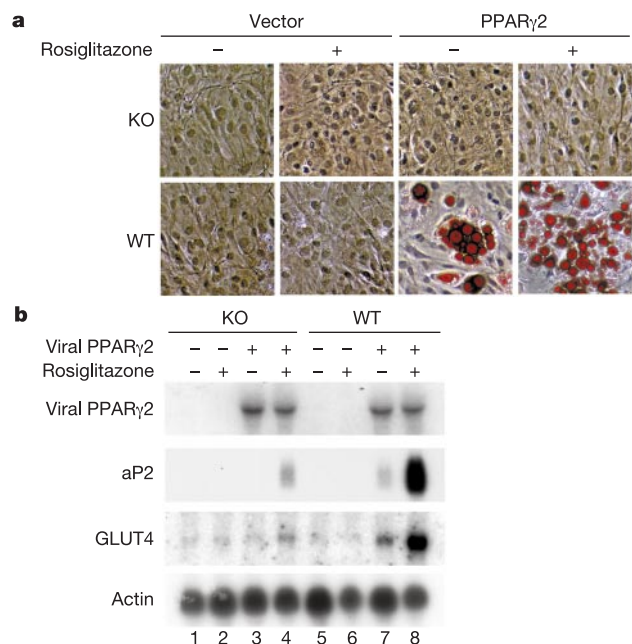


Figure 1 *Trap220*^{−/−} MEFs have defects in PPAR γ 2-stimulated adipogenesis. Cells infected with either a control retroviral vector (pMSCVpuro) or one expressing PPAR γ 2 (pMSCV-PPAR γ 2) were induced with differentiation medium in the presence or absence of 0.5 μ M synthetic PPAR γ ligand rosiglitazone. At day 8 post-induction, cells were either stained for lipid droplets with Oil Red O or total RNA was extracted and subjected to northern blot. **a**, Morphological differentiation shown by Oil Red O staining. **b**, Northern blot analysis of gene expression pattern. KO, *Trap220*^{−/−} cells; WT, wild-type cells.

[‡] Present address: Bristol-Myers Squibb, Wilmington, Delaware 19803, USA.

gene expression.

TRAP220 null mouse embryos die at about embryonic day 11 (E11) owing to developmental defects^{20,21}. Wild-type and *Trap220*^{-/-} MEFs were isolated from E10.0 littermate embryos and self-immortalized²² (see Methods). Immortal cell lines were infected with a retroviral vector expressing PPAR γ 2 (pMSCV-PPAR γ 2). After selection with puromycin, cells were induced with differentiation medium containing the synthetic PPAR γ ligand rosiglitazone or vehicle (dimethyl sulphoxide, DMSO) alone¹⁰. Consistent with previous results¹⁰, expression of PPAR γ 2 induced about 30–40% of wild-type cells to undergo adipogenesis, presumably through the function of endogenous PPAR γ ligand (Fig. 1a, lower panel). Rosiglitazone further stimulated differentiation, resulting in conversion of over 90% of PPAR γ 2-expressing wild-type cells to lipid-laden adipocytes (Fig. 1a, lower panel). In contrast, ectopic expression of PPAR γ 2 failed to induce adipogenesis in *Trap220*^{-/-} cells even in the presence of rosiglitazone, suggesting that *Trap220*^{-/-} cells have defects in PPAR γ 2-stimulated adipogenesis (Fig. 1a, upper panel). Ectopic PPAR γ 2 was expressed at similar levels in wild-type and *Trap220*^{-/-} cells (Fig. 1b, lanes 7 and 8 versus 3 and 4). Consistent with the morphological differentiation, expression of PPAR γ 2 target and adipocyte marker genes such as *aP2* and insulin-responsive glucose transporter (*GLUT4*) were induced in a PPAR γ 2- and TRAP220-dependent manner in both the absence (lane 7 versus lane 3) and, especially, the presence (lane 8 versus lane 4) of ligand. The adipogenesis defects in *Trap220*^{-/-} cells were not aberrant results of cell immortalization because primary *Trap220*^{-/-} MEFs also showed defects in adipogenesis (K.G. and R.G.R., data not shown).

To further verify the role of TRAP220 in adipogenesis, *Trap220*^{-/-} cells were infected with a retroviral vector expressing haemagglutinin- (HA-) tagged TRAP220 (pWZLhygro-TRAP220). After selection with hygromycin, cells were infected with pMSCV-PPAR γ 2 and subsequently induced to differentiate as in Fig. 1. Western blot analysis with anti-HA antibody detected a single band at a relative molecular mass of about 220,000 (*M_r* 220K) only in cells infected with retroviruses carrying TRAP220 (Fig. 2b). Ectopic expression of TRAP220 rescued the adipogenesis defects in *Trap220*^{-/-} cells, with about 70–80% differentiation into adipocytes in the presence of rosiglitazone (Fig. 2a). Consistent with the morphological differentiation, ectopic expression of TRAP220 also rescued the expression of adipogenesis markers that included endogenous PPAR γ , *aP2*, C/EBP α , *GLUT4*, adipin and fatty acid synthase (*FAS*) (Fig. 2c). These data indicate that the adipogenesis defects in *Trap220*^{-/-} cells are due to the absence of the TRAP220 gene, and that TRAP220 is required for PPAR γ 2-stimulated adipogenesis and corresponding target gene expression.

Previous loss-of-function experiments^{6,7}, along with related studies, suggest the existence of a transcriptional cascade in which PPAR γ is downstream of C/EBP β and C/EBP δ but upstream of multiple adipogenesis marker gene expression and terminal differentiation events^{8,23}. Thus, C/EBP β appears to act, at least in part, through PPAR γ and, accordingly, adipogenesis can also be induced by ectopic expression of C/EBP β ²³. We therefore asked whether TRAP220 is also required for C/EBP β -stimulated adipogenesis. Wild-type and *Trap220*^{-/-} cells were infected with a retrovirus vector expressing C/EBP β (pMSCV-C/EBP β). Cells were induced with differentiation medium without PPAR γ ligand. Although the expression levels of C/EBP β before induction were similar among these cells (Fig. 3a), *Trap220*^{-/-} cells almost completely lost the ability to accumulate lipid droplets (Fig. 3b, left versus right panel) or to express adipogenesis markers (Fig. 3c, lane 4 versus lane 6) when compared to wild-type cells. However, the defects in C/EBP β -stimulated adipogenesis in *Trap220*^{-/-} cells could be rescued, at least partially, by ectopic expression of TRAP220 (Fig. 3b, middle panel and Fig. 3c, lane 5). TRAP220 also appears to be required for the expression of endogenous PPAR γ (Fig. 3c, top panel),

suggesting that TRAP220 may be involved not only in expression of PPAR γ target genes but also in expression of PPAR γ itself during C/EBP β -stimulated adipogenesis.

MEFs have the capacity to differentiate into a variety of cell types in addition to adipocytes, as exemplified by ectopic MyoD-induced myogenesis²⁴. To investigate the role of TRAP220 in MyoD-stimulated myogenesis, and also to determine whether *Trap220*^{-/-} cells display a universal defect in various differentiation processes, cells were infected with a retrovirus vector expressing MyoD (pMSCV-MyoD). Three days after induction, about 20% of the *Trap220*^{-/-} cells were identified as myocytes with an elongated and sometimes multi-nucleated phenotype (Fig. 3d, left panel). These myocytes stained positively with anti-MHC (myosin heavy chain) and anti-myogenin antibodies (K.G. and R.G.R., data not shown). Consistent with the morphological differentiation, the expression of muscle-specific MHC and myogenin proteins was induced in a MyoD-dependent manner in *Trap220*^{-/-} cells (Fig. 3e, lane 2 versus 1), suggesting that TRAP220 is not essential for MyoD-stimulated myogenesis. Wild-type cells showed a significantly higher level of myogenesis (70–80% differentiation) than did *Trap220*^{-/-} cells (Fig. 3d, right panel). However, unlike the case in PPAR γ 2- or C/EBP β -stimulated adipogenesis, the relatively low level of myogenesis in *Trap220*^{-/-} cells could not be increased by ectopic expression of TRAP220 (Fig. 3d, middle versus left panels; Fig. 3e, lane 3 versus lanes 1 and 2). This indicates that TRAP220 is not directly involved in MyoD-stimulated myogenesis. The relatively poor myogenic capability of *Trap220*^{-/-} cells may result from developmental defects of the *Trap220*^{-/-} embryos from which

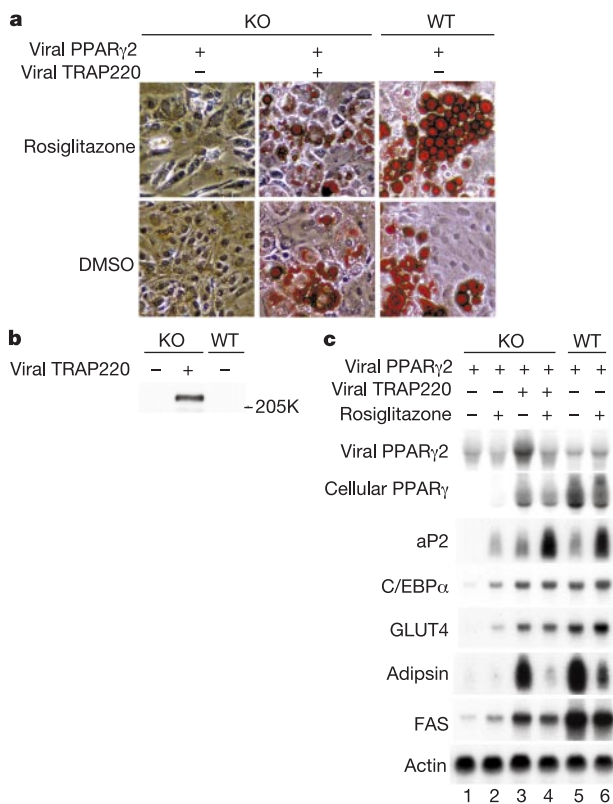


Figure 2 The adipogenesis defects in *Trap220*^{-/-} MEFs can be rescued by ectopic expression of TRAP220. *Trap220*^{-/-} cells infected with either a control retroviral vector (pWZLhygro) or one expressing HA-tagged TRAP220 (pWZLhygro-TRAP220) were selected with hygromycin for two weeks. Cells were subsequently infected with retroviruses expressing PPAR γ 2 and induced to differentiate as in Fig. 1. **a**, Morphological differentiation (Oil Red O staining). **b**, Western blot analysis of ectopic TRAP220 expression. Anti-HA polyclonal antibody was used to detect HA-tagged ectopic TRAP220. **c**, Northern blot analysis of gene expression pattern.

these cells were isolated, or from phenotypic changes that occurred during self-immortalization. Thus, these results show that TRAP220 is selectively required for PPAR γ 2- and C/EBP β -stimulated adipogenesis, relative to MyoD-stimulated myogenesis, and indicate the differential involvement of TRAP220 and the TRAP complex in different fibroblast differentiation pathways.

The demonstration of a ligand-enhanced interaction of PPAR γ 2 with the isolated TRAP220 subunit of the TRAP complex^{3,4} suggested that the latter complex may indeed serve as a coactivator for PPAR γ 2. Consistent with this notion, PPAR γ 2 was shown to interact directly with purified TRAP complex in a ligand-dependent manner (Fig. 4a) and, either directly or indirectly, with TRAP complex in a nuclear extract²⁵ (Fig. 4c, lane 4). Most significantly, however, in an *in vitro* transcription assay reconstituted with highly purified factors, TRAP complex strongly increased the transcriptional activity of PPAR γ 2 on a DNA template containing three copies of the DR1 PPAR γ recognition site (Fig. 4b, lane 5 versus 8). Transcription in the assay was strictly dependent both on PPAR γ 2 and on its DNA binding partner RXR α (Fig. 4b, lane 8 versus lanes 6 and 7). These data show, we believe for the first time, that purified TRAP complex both interacts directly with PPAR γ 2 and functions directly as a transcription coactivator for PPAR γ 2.

A further analysis of the role of TRAP220 in facilitating functional interactions between PPAR γ 2 and the multisubunit TRAP complex, as well as the mechanistic basis for the inability of *Trap220*^{-/-} MEFs to undergo PPAR γ 2-stimulated adipogenesis, showed that TRAP complex from wild-type MEF nuclear extracts bound to GST-PPAR γ 2 whereas TRAP complex from *Trap220*^{-/-} MEF nuclear extracts did not (Fig. 4c, lanes 4 and 5 and data not shown). However, ectopic expression of TRAP220 in *Trap220*^{-/-}

cells restored the ability of TRAP complex in derived nuclear extracts to bind to GST-PPAR γ 2 (Fig. 4c, lane 6). Thus, TRAP220 serves as an essential bridge for the interaction between PPAR γ 2 and TRAP complex. Deletion of TRAP220 disrupts this interaction and probably abrogates the ability of PPAR γ 2 to activate expression of direct target genes that are important for adipogenesis, as shown in the northern blot analyses in Figs 1b and 2c.

This mechanism probably accounts, at least in part, for the inability of *Trap220*^{-/-} cells to undergo PPAR γ 2-stimulated adipogenesis. However, our study has shown that TRAP220 is also required for the activation of PPAR γ expression during C/EBP β -stimulated adipogenesis. This could reflect an essential role for TRAP220 in C/EBP β -mediated activation of PPAR γ expression, in the PPAR γ -C/EBP α cross-regulation loop^{12,26}, or in the function of other transcription factors during adipogenesis. However, our assays to date have failed to show interactions of intact TRAP220 with other factors (such as glucocorticoid receptor, C/EBP α , C/EBP β , SREBP1/ADD1) important for adipogenesis (K.G. and R.G.R., data not shown).

This report provides evidence that the TRAP complex directly facilitates PPAR γ 2-mediated transcription, and that the PPAR γ 2-interacting TRAP220 subunit of the complex is essential both for PPAR γ 2-TRAP interaction and for PPAR γ 2-mediated adipogenesis. The study also provides a demonstration of the involvement of TRAP complex in a specific cell differentiation and gene expression program and, further, shows that a component of the complex (TRAP220) is essential for one embryonic fibroblast differentiation pathway (adipogenesis) but not for another (myogenesis). This provides a model for the regulation of cell-specific transcription and differentiation events through distinct TRAP subunits. As the effects

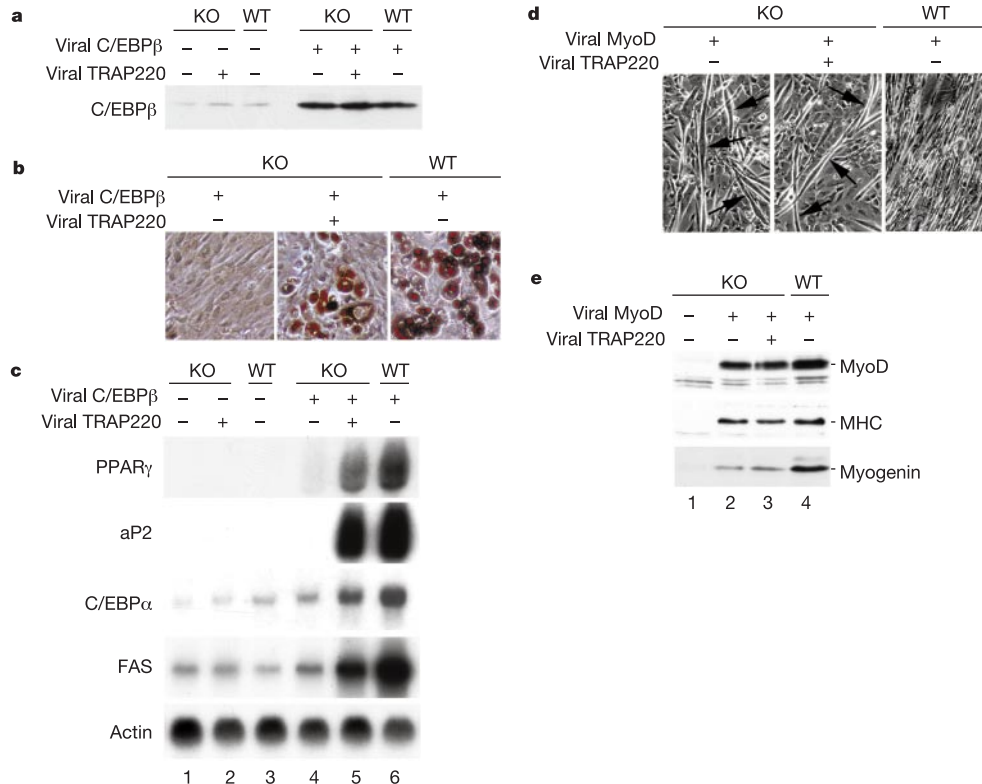


Figure 3 TRAP220 is essential for C/EBP β -stimulated adipogenesis but not for MyoD-stimulated myogenesis. Cells infected with retroviruses expressing either C/EBP β or MyoD were induced to undergo adipogenesis or myogenesis (see Methods). **a**, Expression of C/EBP β protein. Nuclear extracts were prepared from cells before induction of differentiation and subjected to western blot with a polyclonal antibody (Santa Cruz, sc-150) against C/EBP β . **b**, Cell morphology of C/EBP β -stimulated adipogenesis (Oil Red O

staining). **c**, Northern blot analysis of C/EBP β -stimulated adipogenesis. **d**, Phase-contrast microscopy of cells 3 days after induction of MyoD-stimulated myogenesis. Arrows indicate differentiated myocytes. **e**, Expression of MyoD and myogenesis markers. Three days after induction of myogenesis, whole cell lysates were prepared and subjected to western blot with indicated antibodies.

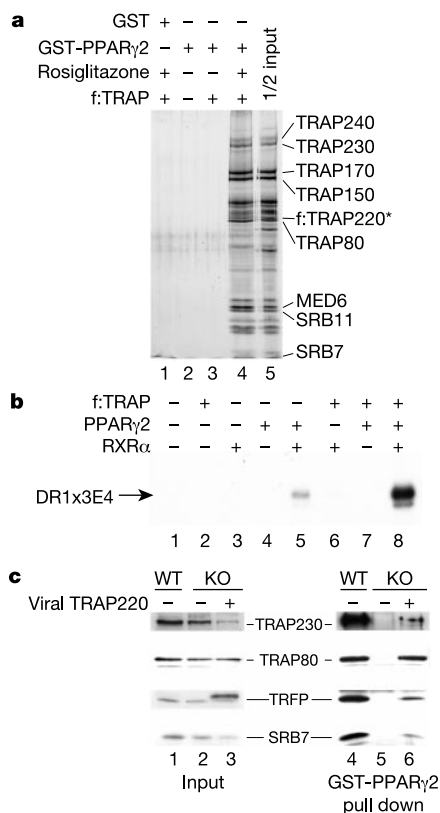


Figure 4 TRAP complex interacts directly with PPAR γ 2, via TRAP220, and is required for optimal PPAR γ 2 function. **a**, Direct interaction between PPAR γ 2 and the TRAP complex. Affinity purified TRAP (f:TRAP) was incubated with immobilized GST or GST-PPAR γ 2 in the presence and absence of ligand and, after washing, bound proteins were analysed by SDS-PAGE with silver staining. **b**, TRAP complex enhances PPAR γ 2/RXR α dependent transcription in a cell free system reconstituted with purified general transcription factors and cofactors (see Methods). **c**, TRAP complex interactions with PPAR γ 2 requires TRAP220. GST-PPAR γ 2 (40 μ g) and 10 μ M rosiglitazone were incubated with nuclear extracts (2 mg protein) prepared from wild-type cells and from *Trap220*^{-/-} (KO) cells with or without retrovirus-mediated expression of TRAP220 as indicated. Bound TRAP complex was monitored by western blot with antibodies to indicated subunits.

of the TRAP complex are apparent on DNA templates, we propose that this complex (and the associated TRAP220) functions downstream of other PPAR γ 2-interacting cofactors (such as PGC-1, CBP/p300 and SRC-1) that have been suggested to act, at least in part, through chromatin modification²⁷. □

Methods

Retrovirus vectors

The pMSCV-PPAR γ 2 vector was prepared by inserting mouse PPAR γ 2 complementary DNA from pBabepuro-PPAR γ 2¹⁰ into pMSCVpuro (Clontech). The 4.7-kilobase human TRAP220 cDNA was modified to express N-terminal Flag and HA tags and cloned into pWZLhygro vector. The pMSCV-C/EBP β vector was constructed by subcloning mouse C/EBP β cDNA from pWZLhygro-C/EBP β into pMSCVpuro. The mouse MyoD cDNA was amplified by polymerase chain reaction (PCR) from pBabe-MyoD²⁴ and inserted into pMSCVpuro. All plasmids were verified by DNA sequencing.

Establishment of wild-type and *Trap220*^{-/-} cell lines

MEF cell lines were self-immortalized following the 3T9 protocol²². Briefly, primary wild-type and *Trap220*^{-/-} MEFs were isolated from E10.0 littermate embryos and cultured in DMEM plus 10% fetal bovine serum. Early passage (less than 5) MEFs were then plated at a density of 2.7×10^6 cells per 10-cm dish. Every 3 days, cells were gently trypsinized and re-plated at the same density. Cells were immortalized after 5 months of continuous culture.

Cell culture, retrovirus infection and induction of differentiation

Wild-type and *Trap220*^{-/-} MEFs, as well as the Phoenix ecotropic retrovirus packaging cell line, were cultured in DMEM plus 10% fetal bovine serum. Recombinant retroviruses were produced as described²⁸. Briefly, Phoenix cells were seeded at a density of

3.2×10^6 cells per 10-cm dish. After 24 h, cells were transfected with 22 μ g of retrovirus vector by lipofectAmine 2000 reagent (Invitrogen). Forty-eight hours after transfection, the cells were changed to fresh medium. After a further 12 h, virus-containing supernatant was collected, filtered through 0.45- μ m membrane and supplemented with 8 μ g ml⁻¹ polybrene (Sigma). MEFs were plated at a density of 5×10^5 cells per 10-cm dish 1 day before infection. Cells were incubated with retrovirus for 36 h and split at a density of 5×10^5 cells per 10-cm dish. Cells were selected with either 2 μ g ml⁻¹ puromycin (Sigma) for 5 days or 150 μ g ml⁻¹ hygromycin (Invitrogen) for two weeks. Induction of adipogenesis was performed as described¹⁰. Briefly, two days after cells reached confluence, cells were treated with medium containing 0.5 mM 3-isobutyl-1-methyl-xanthine (Sigma), 1 μ M dexamethasone (Sigma), 5 μ g ml⁻¹ insulin (Sigma), and 0.5 μ M synthetic PPAR γ ligand rosiglitazone or vehicle (DMSO) alone. Forty-eight hours later, cells were changed to medium containing 5 μ g ml⁻¹ insulin and 0.5 μ M rosiglitazone or DMSO. The medium was replenished at 2-day intervals. At day 8 post-induction, cells were either stained for lipid droplets with Oil Red O (Sigma) or total RNA was extracted and subjected to northern blot analysis. For the myogenesis assay, cells were infected and selected in growth medium (DMEM plus 20% fetal bovine serum). After puromycin selection, cells were grown to 80% confluence and switched to differentiation medium (DMEM containing 2% horse serum and 10 μ g ml⁻¹ insulin) for 3–4 days.

Northern and western blot analyses

Total RNA isolated by TRIzol reagent (Life Technology) was subjected to northern blot analysis as described previously²⁹. Ten micrograms total RNA was loaded in each lane. Whole cell protein extracts prepared as described²⁹ were used for western blot analyses of HA-tagged TRAP220, MyoD, MHC and myogenin. Rabbit anti-HA antibody (Santa Cruz, sc-805) was used at 1:400 dilution. Rabbit anti-MyoD, monoclonal anti-MHC (MF20) and monoclonal anti-myogenin (F5D) antibodies were used as described²⁴.

GST pull down assay

Forty micrograms immobilized GST-PPAR γ 2 (ref. 4) was incubated with rotation for 6 h at 4°C with 2 mg nuclear extracts in BC180/0.1% NP40 and 1 mM dithiothreitol (DTT) in the presence of 10 μ M rosiglitazone. The beads were washed six times with binding buffer, followed by one wash with BC100/0.1% NP40. Bound proteins were eluted by incubating beads twice with BC100/0.1% NP40 plus 0.2% Sarkosyl for 1 h at 4°C and subjected to western blot analysis. To test direct interaction between PPAR γ 2 and TRAP complex, 2.5 μ g immobilized GST-PPAR γ 2 was incubated at 4°C with 500 ng TRAP complex (f:TRAP) affinity purified from HeLa S cell line expressing a Flag-tagged TRAP220 truncation (C.X.Y. and R.G.R., unpublished data). After 2 h incubation in BC180/0.1% NP40 and 1 mM DTT in the presence or absence of 10 μ M rosiglitazone, beads were washed 6 times with binding buffer. Bound proteins were eluted with a synthetic peptide corresponding to the second NR box in TRAP220, resolved in SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and visualized with silver staining.

In vitro transcription assay

These used a highly purified system composed of RNA polymerase II, TFIID, TFIIB, TFIIE, TFIIH and PC4 essentially as described³⁰. Flag-tagged PPAR γ 2, as well as RXR α , were baculovirus expressed and affinity purified. The TRAP complex (f:TRAP) was affinity purified as described above. The naked DNA template (DR1x3E4) contains 3 copies of a DR1 PPAR γ recognition site in front of the adenovirus E4 promoter. Assays were performed in the presence of 10 μ M rosiglitazone. Transcription products were analysed by primer extension.

Received 2 January; accepted 19 March 2002.

- Fondell, J. D., Ge, H. & Roeder, R. G. Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex. *Proc. Natl Acad. Sci. USA* **93**, 8329–8333 (1996).
- Malik, S. & Roeder, R. G. Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells. *Trends Biochem. Sci.* **25**, 277–283 (2000).
- Zhu, Y., Qi, C., Jain, S., Rao, M. S. & Reddy, J. K. Isolation and characterization of PBP, a protein that interacts with peroxisome proliferator-activated receptor. *J. Biol. Chem.* **272**, 25500–25506 (1997).
- Yuan, C. X., Ito, M., Fondell, J. D., Fu, Z. Y. & Roeder, R. G. The TRAP220 component of a thyroid hormone receptor-associated protein (TRAP) coactivator complex interacts directly with nuclear receptors in a ligand-dependent fashion. *Proc. Natl Acad. Sci. USA* **95**, 7939–7944 (1998).
- Barak, Y. et al. PPAR γ is required for placental, cardiac, and adipose tissue development. *Mol. Cell* **4**, 585–595 (1999).
- Kubota, N. et al. PPAR γ mediates high-fat diet-induced adipocyte hypertrophy and insulin resistance. *Mol. Cell* **4**, 597–609 (1999).
- Rosen, E. D. et al. PPAR γ is required for the differentiation of adipose tissue in vivo and in vitro. *Mol. Cell* **4**, 611–617 (1999).
- Lowell, B. B. PPAR γ : an essential regulator of adipogenesis and modulator of fat cell function. *Cell* **99**, 239–242 (1999).
- Tontonoz, P., Hu, E. & Spiegelman, B. M. Stimulation of adipogenesis in fibroblasts by PPAR γ 2, a lipid-activated transcription factor. *Cell* **79**, 1147–1156 (1994).
- Wu, Z. et al. Cross-regulation of C/EBP α and PPAR γ controls the transcriptional pathway of adipogenesis and insulin sensitivity. *Mol. Cell* **3**, 151–158 (1999).
- Glass, C. K. & Rosenfeld, M. G. The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev.* **15**, 121–141 (2000).
- Gu, W. et al. A novel human SRB/MED-containing cofactor complex, SMCC, involved in transcription regulation. *Mol. Cell* **3**, 97–108 (1999).
- Ito, M. et al. Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. *Mol. Cell* **3**, 361–370 (1999).
- Jiang, Y. W. et al. Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. *Proc. Natl Acad. Sci. USA* **95**, 8538–8543 (1998).
- Malik, S., Gu, W., Wu, W., Qin, J. & Roeder, R. G. The USA-derived transcriptional coactivator PC2 is a submodule of TRAP/SMCC and acts synergistically with other PCs. *Mol. Cell* **5**, 753–760 (2000).

16. Rachez, C. *et al.* Ligand-dependent transcription activation by nuclear receptors requires the DRIP complex. *Nature* **398**, 824–828 (1999).
17. Sun, X. *et al.* NAT, a human complex containing Srb polypeptides that functions as a negative regulator of activated transcription. *Mol. Cell* **2**, 213–222 (1998).
18. Naar, A. M. *et al.* Composite co-activator ARC mediates chromatin-directed transcriptional activation. *Nature* **398**, 828–832 (1999).
19. Boyer, T. G., Martin, M. E., Lees, E., Ricciardi, R. P. & Berk, A. J. Mammalian Srb/Mediator complex is targeted by adenovirus E1A protein. *Nature* **399**, 276–279 (1999).
20. Ito, M., Yuan, C.-X., Okano, H. J., Darnell, R. B. & Roeder, R. G. Involvement of the TRAP220 component of the TRAP/SMCC coactivator complex in embryonic development and thyroid hormone action. *Mol. Cell* **5**, 683–693 (2000).
21. Zhu, Y. *et al.* Deletion of PBP/PPARBP, the gene for nuclear receptor coactivator peroxisome proliferator-activated receptor-binding protein, results in embryonic lethality. *J. Biol. Chem.* **275**, 14779–14782 (2000).
22. Todaro, G. J. & Green, H. Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. *J. Cell Biol.* **17**, 299–313 (1963).
23. Yeh, W. C., Cao, Z., Classon, M. & McKnight, S. L. Cascade regulation of terminal adipocyte differentiation by three members of the C/EBP family of leucine zipper proteins. *Genes Dev.* **9**, 168–181 (1995).
24. Novitsch, B. G., Mulligan, G. J., Jacks, T. & Lassar, A. B. Skeletal muscle cells lacking the retinoblastoma protein display defects in muscle gene expression and accumulate in S and G2 phases of the cell cycle. *J. Cell Biol.* **135**, 441–456 (1996).
25. Yang, W., Rachez, C. & Freedman, L. P. Discrete roles for peroxisome proliferator-activated receptor γ and retinoid X receptor in recruiting nuclear receptor coactivators. *Mol. Cell Biol.* **20**, 8008–8017 (2000).
26. Rosen, E. D. *et al.* C/EBP α induces adipogenesis through PPAR γ : a unified pathway. *Genes Dev.* **16**, 22–26 (2002).
27. Puigserver, P. *et al.* Activation of PPAR γ coactivator-1 through transcription factor docking. *Science* **286**, 1368–1371 (1999).
28. Pear, W. S., Nolan, G. P., Scott, M. L. & Baltimore, D. Production of high-titer helper-free retroviruses by transient transfection. *Proc. Natl Acad. Sci. USA* **90**, 8392–8396 (1993).
29. Ge, K. *et al.* Mechanism for elimination of a tumour suppressor: aberrant splicing of a brain-specific exon causes loss of function of Bin1 in melanoma. *Proc. Natl Acad. Sci. USA* **96**, 9689–9694 (1999).
30. Guermah, M., Tao, Y. & Roeder, R. G. Positive and negative TAF(II) functions that suggest a dynamic TFIID structure and elicit synergy with traps in activator-induced transcription. *Mol. Cell Biol.* **21**, 6882–6894 (2001).

Acknowledgements

We thank S.O. Freytag for C/EBP β cDNA and pWZLhygro plasmid; A.B. Lassar for pBabe-MyoD plasmid, anti-myogenin, anti-MyoD and anti-MHC antibodies; T.M. Wilson and S.A. Kliever for rosiglitazone; G.P. Nolan for Phoenix retrovirus packaging cell lines; J. Zhang, S. Malik, U. Kim, X. Ren and other members of the laboratory for critical reading of the manuscript and discussion. This work was supported by the NIH. K.G. was supported by a postdoctoral fellowship from Charles H. Revson Foundation. A.E.W. was supported by the Swedish Cancer Society.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.G.R. (e-mail: roeder@mail.rockefeller.edu).

Functional relationship of cytochrome c_6 and plastocyanin in *Arabidopsis*

Rajeev Gupta, Zengyong He & Sheng Luan

Department of Plant and Microbial Biology, University of California, Berkeley, California 94720, USA

Photosynthetic electron carriers are important in converting light energy into chemical energy in green plants. Although protein components in the electron transport chain are largely conserved among plants, algae and prokaryotes, there is thought to be a major difference concerning a soluble protein in the thylakoid lumen. In cyanobacteria and eukaryotic algae, both plastocyanin and cytochrome c_6 mediate electron transfer from cytochrome b_6f complex to photosystem I^{1–4}. In contrast, only plastocyanin has been found to play the same role in higher

plants. It is widely accepted that cytochrome c_6 has been evolutionarily eliminated from higher-plant chloroplasts^{5,6}. Here we report characterization of a cytochrome c_6 -like protein from *Arabidopsis* (referred to as Atc6). Atc6 is a functional cytochrome c localized in the thylakoid lumen. Electron transport reconstruction assay showed that Atc6 replaced plastocyanin in the photosynthetic electron transport process. Genetic analysis demonstrated that neither plastocyanin nor Atc6 was absolutely essential for *Arabidopsis* growth and development. However, plants lacking both plastocyanin and Atc6 did not survive.

The protein cytochrome c_6 (cyt c_6) was identified from algae more than four decades ago⁷. However, a higher-plant cyt c_6 has never been found, contributing to the dogma that the copper protein plastocyanin is the only electron donor to photosystem I in higher plants⁸. We identified an *Arabidopsis* cyt c_6 -like protein during a yeast two-hybrid screening experiment searching for interacting partners of a chloroplast immunophilin protein. One of the putative interacting clones showed sequence homology to cyt c_6 from cyanobacteria and algae. Southern blot analysis indicated that the complementary DNA came from the *Arabidopsis* genome (data not shown). More recently, the *Arabidopsis* genomic sequence database also showed a single copy of a gene that corresponds to the sequence of our cDNA. Figure 1 presents the cDNA-deduced amino-acid sequence and its alignment with cyt c_6 proteins from cyanobacteria and algae. Although the overall sequence homology is low (20–30% identical between Atc6 and other cyt c_6 proteins), the amino-acid residues for haem-binding are completely conserved. These residues are located in the same relative positions in the Atc6 protein, suggesting that Atc6 may be assembled with a haem group, like other c -type cytochromes (Fig. 1a). Further database searches identified several expressed sequence tags (ESTs) that encode highly similar proteins from other higher plants including soybean, maize and rice (Fig. 1b). Clearly, genes coding for cyt c_6 -like proteins are universally present in higher plants. Identification of ESTs in various plant species and isolation of Atc6 cDNA indicate that genes coding for cyt c_6 are not pseudogenes but are expressed in these plants. We noted an interesting insertion of 12 amino acids in the Atc6 sequence (and other higher-plant sequences) as compared to cyt c_6 proteins from cyanobacteria and algae. This insert contains two cysteine residues that may regulate the function of the protein.

The cDNA-deduced peptide sequence of Atc6 contains a long stretch of amino acids at the amino terminus of the protein that does not align with cyt c_6 proteins from cyanobacteria or algae (Fig. 1a and b). This N-terminal extension may serve as a signal peptide for targeting to a specific subcellular compartment in plant cells. Because cyt c_6 in eukaryotic algae is located in the thylakoid lumen of the chloroplasts, we tested the possibility of Atc6 being located in the thylakoid lumen of chloroplasts in *Arabidopsis* plants. Almost all genes coding for chloroplast proteins are expressed in a green-tissue-specific manner⁹. To determine whether Atc6 gene expression follows this pattern, we isolated total RNA from leaves and roots and analysed the Atc6 messenger RNA levels by northern blot. Figure 2a indicates that Atc6 transcript accumulated in leaves but was not detectable in the roots. We produced recombinant Atc6 protein in *Escherichia coli* as a glutathione S-transferase fusion protein, and used it as antigen to raise antibody in a rabbit. Using a purified antibody, we performed western blot analysis of the Atc6 protein in *Arabidopsis* plants. Figure 2a shows that Atc6 protein, like its mRNA, was produced in leaves but little was detected in the roots.

To facilitate further biochemical studies of Atc6, we overexpressed Atc6 precursor protein under the control of a constitutive promoter in transgenic plants. Atc6 expression was examined at both mRNA and protein level (Fig. 2b). Interestingly, Atc6 mRNA accumulated at high levels in both roots and leaves, but the Atc6 protein was only produced in the leaves. This finding suggests that Atc6 mRNA was not translated, or that the protein was degraded in the roots. Moreover, Atc6 protein in the leaves of both wild-type and

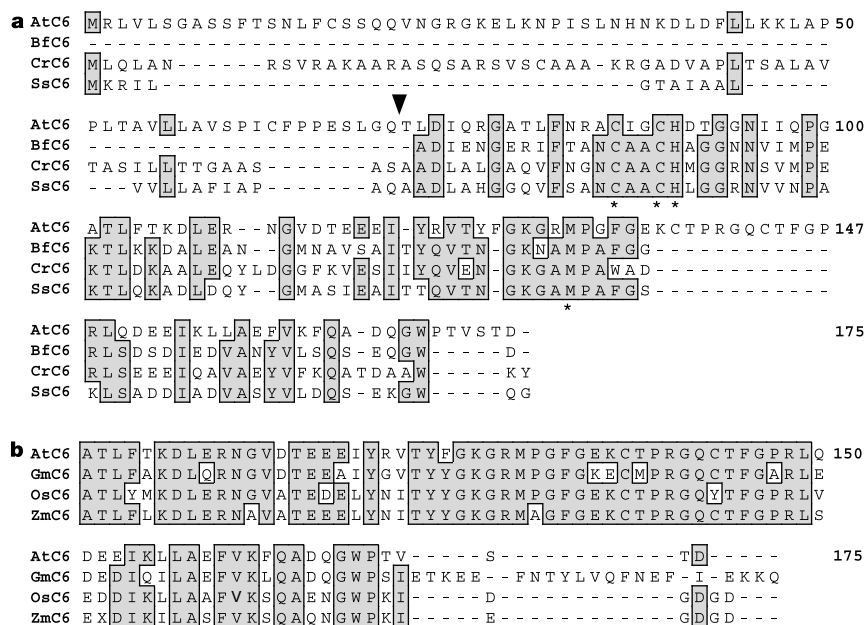


Figure 1 Sequence analyses of Atc6 and other cyt *c*₆ proteins. **a**, Amino-acid sequence alignment of Atc6 and cyt *c*₆ proteins from yellow-green alga (Bf; *Bumilleriopsis filiformis*), green alga (Cr; *Chlamydomonas reinhardtii*) and cyanobacteria (Ss; *Synechococcus* sp. PCC 7942). The haem-binding residues are noted by asterisks.

b, Amino-acid sequence alignment of Atc6 and putative proteins encoded by ESTs from soybean (Gm), maize (Zm) and rice (Os). The vertical arrow in **a** indicates the putative start of mature Atc6 protein. Shaded areas in **a** and **b** are identical amino-acid residues, and dashes represent gaps introduced to optimize alignments.

transgenic plants was processed into a smaller form (relative molecular mass 11,000, *M*_r 11K) as compared to the cDNA-deduced protein (19K), suggesting that Atc6 protein was processed after translation and possible translocation. Post-translational processing and leaf-specific accumulation of Atc6 suggest that this protein may be located in the chloroplast. We purified chloroplasts from both wild-type and cyt *c*₆-overexpressing plants and further fractionated chloroplasts into stromal and thylakoid lumen fractions. Western blot analysis showed that Atc6 protein, like plastocyanin, is located in the thylakoid lumen of both wild-type (not shown) and cyt *c*₆-overexpressing plants (Fig. 2c).

On the basis of the amino-acid sequence (Fig. 1), Atc6 protein

contains the putative haem-binding residues in the conserved positions. To confirm that Atc6 is a functional *c*-type cytochrome, we expressed Atc6 protein in *Synechocystis* sp. PCC 6803 cells lacking the endogenous gene coding for cyt *c*₆ and examined the recombinant protein by haem stain and absorption spectrum analyses. As described in Methods, we constructed a plasmid that targeted the endogenous cyt *c*₆ gene in *Synechocystis* and replaced it with Atc6 cDNA using homologous recombination¹⁰. After confirmation of the gene replacement using Southern blot analysis (data not shown), we cultured the transformant and the wild-type cells

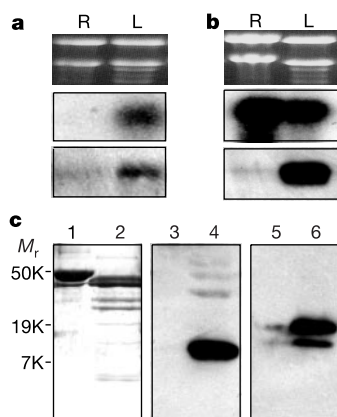


Figure 2 Expression pattern and subcellular localization of Atc6 in *Arabidopsis*. Wild-type *Arabidopsis* (**a**) or transgenic plants overproducing Atc6 (**b**) accumulated Atc6 mRNA (middle panel) and protein (bottom panel). Upper panels in **a** and **b** show ribosomal RNA as a loading control. L, leaves; R, roots. **c**, Localization of Atc6 and plastocyanin. 1, 2, Coomassie-blue-stained proteins from stromal and lumen fraction; 3, 4, a western blot probed with Atc6 antibody against the same proteins as in 1, 2; 5, 6, a western blot probed by a plastocyanin antibody against the same proteins as 3 and 4. The numbers on the left in **c** indicate relative molecular mass.

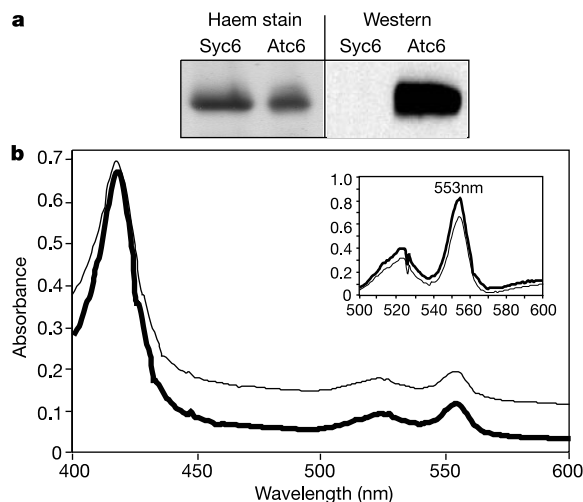


Figure 3 Atc6 is a functional cytochrome *c*. **a**, Haem and antibody stain of *Synechocystis* cyt *c*₆ (Syc6) and Atc6 using Atc6 antibody. **b**, Light absorption analysis of Syc6 and Atc6. For the reduced absorption spectrum, Syc6 and Atc6 was suspended in 50 mM sodium phosphate (pH 7.2) and 100 mM sodium chloride, oxidized by 1.5 mM ferricyanide, and subsequently reduced by 5 mM ascorbate. Reduced minus oxidized absorption spectra of *Synechocystis* cyt *c*₆ and Atc6 are presented in the inset. In both figures, thick and thin lines present Atc6 and *Synechocystis* cyt *c*₆ absorption spectrum, respectively.

under copper-deficient conditions and purified wild-type cyt *c*₆ and Atc6 protein from the cultures. A haem stain assay indicated that both *Synechocystis* cyt *c*₆ and Atc6 proteins are assembled into haem-bound forms (Fig. 3a). In a parallel western blot analysis, *Synechocystis* cyt *c*₆ was not recognized by the Atc6 antibody, consistent with a low sequence similarity between the two proteins. Visible spectral analysis further demonstrated that Atc6 had an absorption spectrum similar to the *Synechocystis* cyt *c*₆ protein (Fig. 3b). The visible absorption spectrum of purified Atc6 in its reduced form exhibited the characteristic absorbency maximum at 553 nm (α), 522 nm (β) and 416 nm (δ or Soret band). The difference spectra (reduced minus oxidized) of Atc6 and Sync6 are also qualitatively identical (Fig. 3b).

A hallmark of cyt *c*₆ function in cyanobacteria and eukaryotic algae is to replace plastocyanin in the photosynthetic electron transfer process. To determine whether Atc6 retains the function of a typical cyt *c*₆ protein, we tested its electron transport property in an oxygen evolution reconstruction assay. As previously demonstrated¹¹, intact chloroplasts or thylakoids are capable of light-dependent oxygen evolution in the presence of a terminal electron acceptor. If inside-out vesicles of thylakoid membrane are used, however, oxygen production is drastically reduced because such preparations largely lack plastocyanin, the lumen electron carrier that is lost during preparation. Addition of plastocyanin or an equivalent electron carrier to the inside-out vesicles restores oxygen evolution¹². We prepared inside-out thylakoid vesicles and used this preparation to test the function of Atc6. As shown in Table 1, intact thylakoids evolved oxygen at high rates as expected. Inside-out vesicles produced much less oxygen (19% of the intact thylakoids). When recombinant *Arabidopsis* plastocyanin (AtPC), *Synechocystis* plastocyanin, *Synechocystis* cyt *c*₆, or Atc6 protein was added to the preparation, oxygen evolution was restored to 89%, 72%, 73% and 74% of the level in the sample containing intact thylakoids. This result demonstrated that Atc6 protein functionally replaces plastocyanin in the photosynthetic electron transport, as a typical cyt *c*₆ protein does in cyanobacteria and algae. Together with the fact that Atc6 protein is located in the thylakoid lumen, these results strongly suggest that Atc6 functions along with plastocyanin in photosynthetic electron transport.

To further determine the physiological function of Atc6 and its relationship with plastocyanin, we used a reverse genetics method to disrupt the function of Atc6. We identified a T-DNA insertional allele of *Atc6* gene from a collection of *Agrobacterium tumefaciens*-DNA-transformed *Arabidopsis* lines (Fig. 4a). The *Agrobacterium* T-DNA was inserted in the second exon of the *Atc6* gene, causing the disruption of the coding sequence at a position corresponding to amino acid residue V59 (Fig. 4a). As a result, the knockout mutant plants did not express Atc6 mRNA or protein (Fig. 4b). The homozygous *Atc6* mutant plants did not show visible phenotypic changes during their life cycle (Fig. 4e and f), which was expected if Atc6 and plastocyanin are functionally equivalent. It became clear that one must examine the function of Atc6 in the plastocyanin-null background, which presented a technical challenge. Plants with plastocyanin-null genetic background have not been generated previously in *Arabidopsis* (or any other plant species). Therefore,

it is not known whether plastocyanin is essential for plant growth and development despite the extensive biochemical and biophysical studies on the function of plastocyanin in the electron transfer process. Without a plastocyanin mutant, it prevented genetic crosses between *Atc6* mutant and plastocyanin-null plants. Although it is theoretically possible to remove functional plastocyanin from plants by copper starvation⁴, it has not been demonstrated in higher plants. Nevertheless, we grew the wild-type and *Atc6* mutant plants in a liquid medium without copper, and monitored the loss of plastocyanin protein and phenotypic changes.

Western blot analysis showed that our procedure reduced the level of plastocyanin protein by 50–60% in both the wild-type and the *Atc6* mutant plants during a two-week culture (data not shown). Reduction in plastocyanin protein abundance did not cause significant phenotypic changes in either wild-type or *Atc6* mutant (data not shown). In addition, reduction in plastocyanin under copper-deficient conditions did not alter the level of Atc6 protein in wild-type plants, suggesting that Atc6 expression is not regulated by plastocyanin abundance in plants as it is in cyanobacteria and algae. Prolonged culture under copper-deficient condition caused

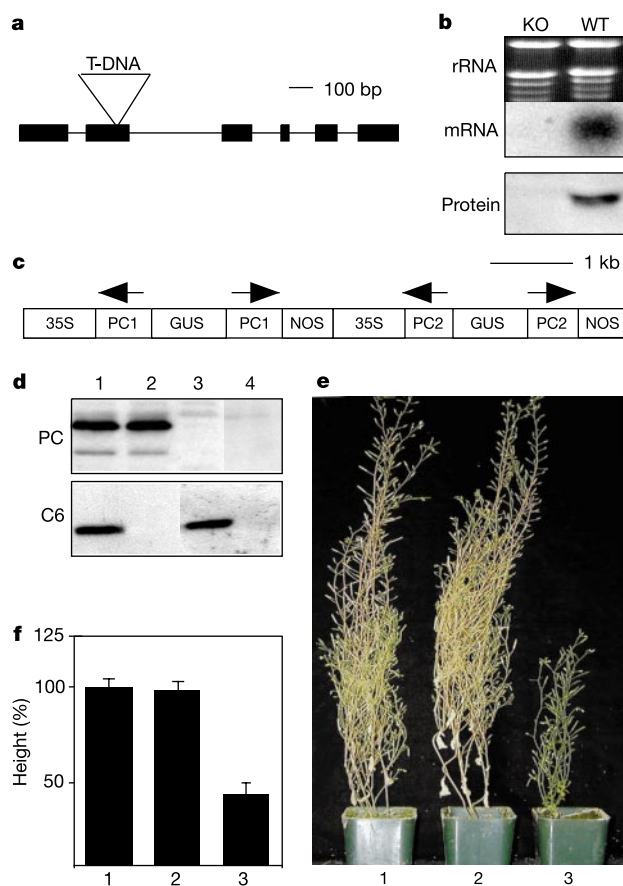


Figure 4 Genetic analysis of plastocyanin and Atc6 in *Arabidopsis*. **a**, T-DNA insertional allele of the *Atc6* gene. Black boxes and lines indicate exons and introns, respectively. The size of T-DNA is not drawn to scale. **b**, Atc6 mRNA and protein in wild-type (WT) and mutant (KO) plants. The rRNA was shown as a loading control for mRNA blot. **c**, PC1/PC2-RNAi construct. Arrows indicate the 5' → 3' direction. Transcription cassette contains partial plastocyanin cDNAs (PC1 and PC2), partial GUS gene (*GUS*), CaMV 35S promoter (35S), and nopaline synthetase terminator (NOS). **d**, Western blot analysis of wild-type (1, ecotype WS), *Atc6* knockout mutant in WS background (2); PC1/PC2-RNAi construct in WS background (3), and PC1/PC2-RNAi construct in *Atc6* mutant background (4). **e**, Phenotype of wild-type (1), *Atc6* mutant (2), and wild-type background without plastocyanin protein (3). **f**, Statistical analysis of plant height of three independent transgenic lines (10 plants in each) in the same genetic background as in **e**. Shown as 100% of wild-type control. The error bars indicate the standard deviation in each sample.

Table 1 Atc6 replaces plastocyanin in the O₂ evolution assay

Assay conditions		O ₂ evolution (μ mol per mg chl)	Percentage
Thylakoids	Buffer	228.25 $\pm$ 23	100 (control)
Inside-out vesicles	Buffer	44.5 $\pm$ 10	19
	Sync6	166.75 $\pm$ 22	73
	Atc6	169 $\pm$ 16	74
	SynPC	165.87 $\pm$ 15	72
	AtPC	202 $\pm$ 11	89

Sync6, *Synechocystis* cyt *c*₆; SynPC, *Synechocystis* plastocyanin; AtPC, *Arabidopsis* plastocyanin.

bleaching and death of plants, consistent with the understanding that copper is an essential microelement for plant growth. However, it is not possible to determine whether plant death was a result of plastocyanin depletion or a defect in other copper-dependent processes.

We examined the functional relationship of Atc6 and plastocyanin by genetically silencing plastocyanin expression in the wild-type and Atc6-null genetic background. The *Arabidopsis* genome contains two genes coding for plastocyanin (*PC1* and *PC2*, Genbank accession numbers P11490 and CAB66894) that are highly homologous to each other at protein level. To disrupt the function of plastocyanin, we must silence both genes. We achieved this goal by the RNA interference (RNAi) approach that has been used to silence genes in both animals and plants^{13,14}. The RNAi construct contained the separate inverted repeats of *PC1* and *PC2* partial cDNA sequences with a partial β -glucuronidase (*GUS*) gene as a spacer in between these repeats (Fig. 4c). The expression of *PC1* and *PC2* RNAi products was driven by a separate CaMV 35S promoter that is highly active in most plant tissues including photosynthetic cells¹⁵. This construct (named as 'PC1/PC2-RNAi') was introduced into both the wild-type (WS ecotype) and the Atc6 mutant to produce transgenic plants. The transgenic wild-type plants with PC1/PC2-RNAi construct produced normal level of Atc6 protein, but plastocyanin was not detectable by western blot analysis (Fig. 4d). The transgenic plants expressing the RNAi construct in the Atc6 mutant background did not express detectable levels of plastocyanin or Atc6 (Fig. 4d). Transgenic plants without detectable plastocyanin protein were significantly smaller than wild-type plants (Fig. 4e, f). Transgenic plants lacking both plastocyanin and Atc6 protein (20 independent lines) did not survive under the same growth condition (not shown). These plants died soon after transfer from the agar plate to the soil. These results from genetic analyses demonstrated that neither plastocyanin nor Atc6 is absolutely essential for plant growth and development. While plastocyanin completely complemented the Atc6 mutant, Atc6 partially replaces plastocyanin under normal growth conditions. Our results also defined another critical difference between higher plants and algae in the regulation of Atc6 expression. The cyt *c*₆ protein is expressed in algae and cyanobacteria only if plastocyanin is depleted under copper deficiency conditions⁴. In higher plants, Atc6 protein is constitutively expressed and is not regulated by plastocyanin abundance.

We have identified a functional cyt *c*₆ from a higher plant. Sequence search results indicated that other plants such as soybean, maize and rice express a similar gene, suggesting that all higher plants possess cyt *c*₆ that is located in the thylakoid lumen. Biochemical and genetic analyses demonstrated the functional relationship of Atc6 and plastocyanin in *Arabidopsis*. This study has added an ancient protein to the list of electron carriers in higher plants, and provided evidence against the accepted view of electron transport in higher plants versus eukaryotic algae and cyanobacteria. □

Methods

Molecular cloning, transgenic plant, and genetics procedures

The Atc6 cDNA was cloned by a yeast two-hybrid screening for interactors of a chloroplast immunophilin protein. The *Arabidopsis* ACT cDNA expression library CD4-22 (ref. 16) was screened essentially as described in ref. 17. For northern blot analysis, *Arabidopsis* plants (ecotype Columbia) were grown in a greenhouse under long-day conditions to flowering stage. Different organs were gathered and total RNA was isolated and analysed as described¹⁸.

To identify a T-DNA insertional allele in the *Atc6* gene, we screened 62,000 T-DNA-transformed *Arabidopsis* lines using polymerase chain reaction (PCR) as previously described¹⁹. The homozygous knockout plants were analysed for Atc6 mRNA and protein levels, and were used for further genetic analyses. To produce transgenic plants overexpressing Atc6, the coding sequence of Atc6 precursor was cloned in sense orientation following a constitutive promoter in the binary vector pATC940²⁰. To disrupt the function of both *PC1* and *PC2* genes, we used the RNAi approach previously described¹⁴ with a modification to silence both genes with one construct. The 'PC1/PC2-RNAi' construct as shown in Fig. 4c was transferred into *Agrobacterium tumefaciens* GV3101pMP9 strain for *Arabidopsis* transformation. *Arabidopsis* plants (ecotype Columbia or WS or Atc6 mutant) were transformed by the 'floral dip' method²¹.

To produce Atc6 antibody, the mature form of Atc6 was expressed in *E. coli* as a glutathione S-transferase (GST) fusion protein that was used as antigen to immunize rabbit. The Atc6 antibody was purified by an affinity purification method²². For western blot analysis, total protein was extracted from plants, separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE), and transferred onto nitrocellulose membrane for antibody probing. The amount of bound antibody was detected by the ECL system (Amersham).

For protein localization assay, chloroplasts were isolated from protoplasts according to a previous method²³ with modifications. Protoplast suspension was passed through 15- μ m nylon net to release chloroplasts. Chloroplasts were lysed on ice in a hypotonic solution. The thylakoids were sonicated to release the lumen contents, and centrifuged to separate the membrane and the soluble fraction. The stromal and luminal fraction was analysed by western blot to determine the distribution of plastocyanin and Atc6 proteins in the two fractions.

Cyanobacterial genetics and biochemical procedures

Wild-type *Synechocystis* sp. PCC 6803 strain was obtained from A. Glazer and grown in BG-11 medium²⁴ at 30 °C. Transformation of *Synechocystis* sp. PCC 6803 strains and mutant selection was carried out according to ref. 10. The mutant segregation was followed by PCR and confirmed by Southern hybridization. The genomic sequence information was obtained from CyanoBase (<http://www.kazusa.or.jp/cyano>).

To express *Arabidopsis* cyt *c*₆ (Atc6) and plastocyanin (AtPC) in *Synechocystis* sp. PCC 6803, the *petJ* and *petE* gene encoding *Synechocystis* cyt *c*₆ (SynC6) and plastocyanin (SynPC) was deleted and replaced by *petJ* and *petE* from *Arabidopsis*. To delete the *petE* gene from *Synechocystis* sp. PCC 6803, a 4-kilobase genomic fragment containing the 0.5-kb *petE* coding region and 1.75 kb upstream and downstream sequences was amplified by PCR and cloned into pUC119. The coding region for the mature protein was deleted by PCR using the QuikChange site-directed mutagenesis kit (Stratagene) and replaced by the kanamycin-resistance cassette from pUC4k (Pharmacia) and used to transform wild-type *Synechocystis* sp. PCC 6803. Kanamycin-resistant transformants were segregated to produce *petE* mutant. A similar strategy was used to create the *petJ* mutant. The *Synechocystis* strains lacking *petE* and *petJ*, respectively, were transformed by plasmids containing *Arabidopsis* plastocyanin and cyt *c*₆ cDNA. These two plasmids were constructed by fusing the cDNA fragment encoding the mature region of AtPC or Atc6 to the genomic sequence for leader peptide of *Synechocystis* sp. PCC 6803 *petE* and *petJ* gene. The chlorophenicol-resistance cassette from pACYC184 was inserted downstream of the stop codon. Transformants were selected on medium containing chlorophenicol and segregated to homogeneity.

Wild-type *Synechocystis* sp. PCC 6803 strain was used to purify SynPC and SynC6 protein. The mutant strains expressing AtPC and Atc6 (as described above) were used to purify *Arabidopsis* proteins. All strains were grown in BG-11 medium with (for plastocyanin) or without (for cyt *c*₆) copper according to previously described procedures^{25,26}. Purified Atc6 was confirmed by western blot, haem stain, and spectrum analysis. The identity and purity of plastocyanin protein was verified by SDS-PAGE and western blot. Haem-dependent peroxidase activity was determined by a procedure described elsewhere²⁷.

Inside-out thylakoid vesicles were prepared from spinach using a previously reported method¹¹. Whole chain electron transport from water to methyl viologen was followed at 25 °C using a Clark-type oxygen electrode. The assay medium contains 20 mM MES, pH 6.5, 100 mM sucrose, 5 mM KCl, 5 mM MgCl₂, 0.1 mM methyl viologen, and intact thylakoids or inside-out vesicles corresponding to 50 μ g chlorophyll in a total volume of 2 ml. Where indicated, plastocyanin or cyt *c*₆ was added to the reactions to a final concentration of 1.5 μ M.

Received 11 January; accepted 25 March 2002.

- Crofts, A. R. & Wood, P. M. Photosynthetic electron transport chains in plants and bacteria and their role as proton pumps. *Curr. Top. Bioenerg.* **7**, 175–244 (1978).
- Hauska, G. A., McCarthy, R. E., Berzborn, R. J., Racker, E. J. *Biol. Chem.* **246**, 3524–3531 (1971).
- Malkin, R. & Niyogi, K. in *Plant Biochemistry and Molecular Biology* (eds Buchanan, B. B., Gruissem, W. & Jones, R. L.) 568–629 (American Society of Plant Physiologists, Rockville, Maryland, 2000).
- Merchant, S. in *Metal Ions in Gene Regulation* (eds Silver, S. & Walden, W.) 450–467 (Chapman and Hall, New York, 1998).
- De la Cerdá, B., Diaz-Quintana, A., Navarro, J. A., Hervás, M. & De la Rosa, M. A. Site-directed mutagenesis of cytochrome c6 from *Synechocystis* sp. PCC6803: The heme protein possesses a negatively charged area that may be isofunctional with the acidic patch of plastocyanin. *J. Biol. Chem.* **274**, 13292–13297 (1999).
- Hippler, M., Drepper, E., Roach, J.-D. & Muehlenhoff, U. Insertion of the N-terminal part of Psaf from *Chlamydomonas reinhardtii* into photosystem I from *Synechococcus elongatus* enables efficient binding of algal plastocyanin and cytochrome c6. *J. Biol. Chem.* **274**, 4180–4188 (1999).
- Katoh, S. Studies on the algal cytochromes of C-type. *Biochem. J.* **62**, 629–632 (1959).
- Sigfridsson, K. Plastocyanin, an electron-transfer protein. *Photosynth. Res.* **57**, 1–28 (1998).
- Tobin, E. M. & Silverthorne, J. Light regulation of gene expression in higher plants. *Annu. Rev. Plant Physiol.* **36**, 569–594 (1985).
- Williams, J. G. Construction of specific mutations in photosystem II photosynthetic reaction center by genetic engineering methods in *Synechocystis* 6803. *Methods Enzymol.* **167**, 766–778 (1988).
- Atta-Asafo-Adjei, E. & Dilley, R. A. Plastocyanin stimulation of whole chain and photosystem I electron transport in inside-out thylakoid vesicles. *Arch. Biochem. Biophys.* **243**, 660–667 (1985).
- Wood, P. M. Interchangeable copper and iron proteins in algal photosynthesis. Studies on plastocyanin and cytochrome c-552 in *Chlamydomonas*. *Eur. J. Biochem.* **87**, 9–19 (1978).
- Waterhouse, P. M., Graham, M. W. & Wang, M. B. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc. Natl Acad. Sci. USA* **95**, 13959–13964 (1998).

14. Chuang, C.-F. & Meyerowitz, E. M. Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **97**, 4985–4990 (2000).
15. Luan, S. & Bogorad, L. A rice gene promoter contains separate cis-acting elements for the expression in dicot and monocot plants. *Plant Cell* **4**, 971–981 (1992).
16. Kim, J., Harter, K. & Theologis, A. Protein-protein interactions among the AUX/IAA proteins. *Proc. Natl Acad. Sci. USA* **94**, 11786–11791 (1997).
17. Durfee, T. *et al.* The retinoblastoma protein associates with the protein phosphatase type-1 catalytic subunit. *Genes Dev.* **7**, 555–569 (1993).
18. Gupta, R., Huang, Y., Kieber, J. J. & Luan, S. Identification of a dual-specificity protein phosphatase that inactivates MAP kinase from *Arabidopsis*. *Plant J.* **16**, 581–590 (1998).
19. Krysan, P. J., Young, J. C. & Sussman, M. R. T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell* **11**, 2283–2290 (1999).
20. Ni, M., Cui, D., Einstein, J., Narasimulu, S., Vergara, C. E. & Gelvin, S. B. Strength and tissue specificity of chimeric promoters derived from the octopine and mannopine synthase genes. *Plant J.* **7**, 661–676 (1995).
21. Clough, S. J. & Bent, A. F. Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
22. Lin, Y., Wang, Y., Zhu, J.-K. & Yang, Z. Localization of a Rho GTPase implies a role in tip growth and movement of the generative cell in pollen tubes. *Plant Cell* **8**, 293–303 (1996).
23. Somerville, C. R., Somerville, S. C. & Ogren, W. L. Isolation of photosynthetically active protoplasts and chloroplasts from *Arabidopsis thaliana*. *Plant Sci. Lett.* **21**, 89–96 (1982).
24. Rippka, R., Deruelles, J., Waterbury, J., Herdman, M. & Stanier, R. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J. Gen. Microbiol.* **111**, 1–61 (1979).
25. Hervas, M. *et al.* *Synechocystis*-6803 plastocyanin isolated from both the cyanobacterium and *E. Coli* transformed cells are identical. *FEBS Lett.* **319**, 257–260 (1993).
26. Zhang, L., McSpadden, B., Pakrasi, H. & Whitmarsh, J. Copper-mediated regulation of cytochrome-c553 and plastocyanin in the cyanobacterium *Synechocystis*-6803. *J. Biol. Chem.* **267**, 19054–19059 (1992).
27. Quinn, J. M. & Merchant, S. Copper-responsive gene expression during adaptation to copper deficiency. *Methods Enzymol.* **297**, 263–279 (1998).

Acknowledgements

We thank R. Malkin, J. Gray and S. Merchant for discussions. This work was supported by the National Institutes of Health and the US Department of Energy (S.L.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.L. (e-mail: sluan@nature.berkeley.edu). The Atc6 cDNA sequence has been deposited at GenBank under accession number AJ438488.

corrigendum

Isochorismate synthase is required to synthesize salicylic acid for plant defence

Mary C. Wildermuth, Julia Dewdney, Gang Wu & Frederick M. Ausubel

Nature **414**, 562–565 (2001).

In this Letter, the GenBank accession number for the sequence of ICS1 cDNA was given incorrectly as AY056500. It should be AY056055. □

Working with genes

Gene expression, analysis, hybridization and delivery.



ABI's PRISM 3100-Avant.

Genetic analyser

Applied Biosystems

<http://home.appliedbiosystems.com>

Entry-level genetic analyser

The ABI PRISM 3100-Avant Genetic Analyser is a four-capillary electrophoresis system for low- to medium-throughput DNA analysis. Increased sensitivity and efficiency of the system is claimed to reduce the amount of DNA required per sample injection, making it extremely cost-effective. The 3100-Avant can be upgraded to 16-capillary ABI PRISM 3100 Genetic Analyser specification. The analyser uses ABI PRISM BigDye Terminator chemistries for sequencing and five-dye chemistries for fragment analysis.

JetPEI transfection reagents

Qbiogene

www.qbiogene.com

Efficient gene delivery with low toxicity

This new line of cationic polyethylenimine (PEI) derived transfection reagents has been developed for gene delivery to various established cell lines and primary cells. An affordable and well-characterized alternative to liposome-based transfection, jetPEI reagents are claimed to provide effective and highly reproducible gene and oligonucleotide delivery with low residual toxicity. The jet-PEI line includes reagents for targeting specific adherent and suspended cell types. jetPEI Transfection Reagents have been used successfully with over three dozen different cell lines.

Human genes

Cambridge BioScience www.bioscience.co.uk

Full-length genes for functional genomics

Cambridge BioScience now offers 10,000 human full-length cDNA clones in the UK and Ireland. The clones are produced by OriGene Technologies and represent about 70 % of the mRNA sequences annotated by the National Centre for Biotechnology Information. OriGene aims to assemble a collection

of full-length cDNA clones for every human gene. To search for a specific gene go to www.bioscience.co.uk and enter the gene's NM accession number or nucleotide sequence, or perform a BLAST search. Alternatively, search for multiple genes with specific structural and functional domains. Plasmid DNA containing an insert of the appropriate cDNA fragment is provided in a dry filter, ready for elution and transformation into a competent bacterial host.

Gene delivery system

Stratagene

www.stratagene.com

No help required

The AAV Helper-Free System is a gene delivery system intended for stable, long-term gene expression. It is based on recombinant adeno-associated virus (AAV) technology but eliminates the need for helper virus. Stratagene says that high-efficiency gene delivery is possible for a broad range of hosts, overcoming the limitations of traditional transfection methods. AAV is naturally replication-deficient and normally requires coinfection with an unrelated helper virus to generate AAV virions. Stratagene's system uses a vector containing the necessary genes from adenovirus (pHelper vector) to induce the lytic phase of AAV producing recombinant, replication-defective AAV virions ready to deliver a gene of interest to target cells.

Nucleon

Tepnel Life Sciences

www.tepnel.com

Easily disposed of

The Nucleon range of DNA purification kits has been extended with the addition of two new chloroform-free kits for the purification of genomic DNA from whole blood and mousetails. To eliminate concerns over the use and disposal of chloroform the new kits use no organic solvents and do not generate any chlorinated waste by-products. As the Nucleon DNA extraction system is not based on a column format, scale-up to large sample volumes is simple and economical.

GeneXPRO Contract Services

InnoGenex

www.innogenex.com

Out with it

GeneXPRO contract services make it possible to outsource gene expression profiling, target localization and analysis of cellular and tissue level expression. The company's gene profiling services include design, synthesis and labelling of the probes, correlation of gene expression profile with normal and patho-

genic conditions on cells, tissues and tissue arrays and determination of changes in gene expression in response to candidate drugs. Assays include *in situ* hybridization to determine gene expression in different cells and tissues at different developmental stages.

Tm Analysis System

Shimadzu

www.shimadzu.com

Thermal melt analysis

Intended for antisense, antigene and hybridization studies, the Shimadzu Scientific Instruments Tm Analysis System complements Shimadzu's range of UV-visible spectrophotometers. The Tm Analysis System monitors nucleic acid dissociation as temperature is raised by following absorbance changes at user-specified wavelengths and determines the melting temperature at which half of the double stranded form separates. The thermodynamics parameters obtained enable the prediction of thermal stability and structure of nucleic acids, target validation by antisense and antigene methods, as well as analysis and optimization of a variety of techniques based on hybridization phenomena.

These notes are compiled in the Nature office from information provided by the manufacturers.

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Laura Pearson (4977)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silke Opstrup (4994)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau
Tel + 33 (0) 1 43 20 16 51
Fax + 33 (0) 1 43 20 51 52
e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan
Tel + 49 89 54 90 57 11/2
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager:

Payton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakatomachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.johnson@naturejpn.com

Asia-Pacific Advertising Manager:

Kevyn Johnson

naturejobs

Taking stock of biotech jobs

One of the most obvious ways to assess the state of the biotechnology job market is to turn to the business news pages. There, the stories paint an almost uniformly glum picture. In Britain, for example, few biotech firms have made independent public offerings over the past 16 months. And in Germany, biotech companies still have not recovered from the crash of the country's junior exchange last year. Meanwhile, many life-sciences funds in the United States have lost 30% of their value since the start of the year.

But curiously, the news does not accurately reflect the biotech recruitment market. For example, both Genentech in South San Francisco and Genzyme in Cambridge, Massachusetts, have had over 30% of their valuation wiped out since December — but their websites are still posting scientific vacancies. Similarly, Icos, based in Bothell, Washington, suffered a major setback when its lead drug — an impotence treatment that would compete with Viagra — was delayed for a year, but it too is still recruiting.

So what effect has the market had on scientific jobs? Perhaps healthier financial results would have led to even greater recruitment drives. But it is more likely that the effects are being felt away from the limelight — by the start-up company that cannot hire more researchers because it needs another infusion of venture capital, for example, or by the mid-sized firm that, without an independent public offering, cannot get a promising compound through clinical trials on its own.

Perhaps the biggest cultural effect is a 'turning of the tables'. It is probably fair to say that academics who stayed put during the biotech boom are now entitled to feel smug, much as their colleagues who left for industry felt when they read the papers back when their stock options looked more promising.

Paul Smaglik

Naturejobs editor



Erratum A caption in the Regions report on Boston/Cambridge in the 16 May issue of *Naturejobs* incorrectly identified AstraZeneca's Boston-area research facility as the Whitehead Institute. *Naturejobs* apologizes for any confusion.

Contents

CAREERS AND RECRUITMENT

Collaborations in climate change p4

The changing climate of Japanese research p7

WWW.NATUREJOBS.COM

Career centre
Information on the scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



Climate of care for a changing world

The Intergovernmental Panel on Climate Change is fostering collaborations and emphasizing input from developing countries, says Heike Langenberg.

Every few years, a flurry of activity sweeps through the climate-research community. Around the world, model simulations are hastily forced through supercomputers, papers are finalized and submitted, and scientists travel from meeting to meeting. Such a commotion indicates the imminent publication of the latest assessment report by the Intergovernmental Panel on Climate Change (IPCC), a group of experts that was founded jointly by the World Meteorological Organization and the United Nations Environment Programme in 1988 to establish an international consensus on climate-change research.

The panel — whose hundreds of members spend nights, weekends and late evenings debating the assessment's wording — has an indisputable policy impact; the world is still responding to its 2001 bombshell: "There is new and stronger evidence that most of the warming observed over the last 50 years is attributable to human activities." But the IPCC process also alters the way climate research proceeds.

PERIODIC CONSENSUS

Now, in the wake of the 2001 report, researchers find themselves back at their benches, trying to fill the gaps that such a comprehensive synthesis invariably reveals. "The IPCC process is very stimulating," says Gerald Meehl of the National Center for Atmospheric Research in Boulder, Colorado, who is a lead author on the chapter covering projections of future climate change. "As part of the process of assessing possible future changes of climate, it became apparent that there are gaps in our knowledge," he says.

The possible changes in the frequency of extreme events is one area identified as lacking in the 2001 report, says Meehl. He hopes that by the time the next set of reports is published, currently scheduled for 2007, some of these gaps will be closed. Two recent papers in *Nature* assessing extreme rainfall seasons and

flooding events throughout the world have made a start on this exploration (see T. N. Palmer & J. Räisänen, *Nature* **415**, 512–514; 2002 and P. C. D. Milly, R. T. Wetherald, K. A. Dunne & T. L. Delworth, *Nature* **415**, 514–517; 2002).

ONE WORLD

The IPCC is a truly global effort, involving representatives from countries all over the world. Most of the research that is assessed comes from the industrialized countries with their larger budgets for climate research and, consequently, greater volume of publications. But each of the three working groups and the task force is either chaired or co-chaired by a scientist from a developing country, and all the smaller groups working on single chapters include researchers from the developing world. This attitude has been underlined this year with the election of

Gerald Meehl aims to fill gaps in current knowledge.



Rajendra Pachauri from India as chairman of the panel — the first IPCC leader to come from a developing country (see *Nature* **416**, 771 & 774; 2002).

“We encourage scientists from developing countries and economies in transition to participate in the IPCC process and also to host meetings,” says Renate Christ, deputy secretary of the IPCC. “It is more difficult to find the right people in countries outside the industrialized nations, but over the past years we have built a network of excellent scientists.”

Meetings are held in places such as Arusha (Tanzania) or Beijing, which might seem remote to scientists from the United States or Europe. Comments like “frequent flyers are very happy” float around in the research community. But locations all over the world have been chosen deliberately to encourage local scientists and students to attend and eventually contribute.

BUILDING BRIDGES

The effort appears to be paying off. Jose Marengo is a coordinator of climate studies at the Centre for Weather Forecast and Climate Studies in Brazil and has been involved in the IPCC since 1994. Born and educated to BSc level in Peru, and with doctoral and postdoctoral training in the United States, he says: “Participation from Peru was very low initially. But when a meeting was held in Lima, I contacted some of my old colleagues and fellow students. Now, the country is well represented.”

Making a similar point, the latest assessment — the Working Group II report on Impacts, Adaptation and Vulnerability — includes regional chapters with a large majority of authors from the region that is assessed: Africa, Asia, Australia and New Zealand, Europe, Latin America, North America, Polar Regions and Small Island States. Attracting scientists from all over the world has obviously been successful, even though the number of climate researchers is small in many countries. “Brazil, with probably fewer than 40 climate researchers, is still one of the largest climate-research communities in South America,” says Marengo.

The lack of researchers working on climate issues can cause problems for the developing world in trying to participate, says Ray Pierrehumbert, professor of geology and geophysics at the University of Chicago, who worked on the chapter on physical processes and feedback. “Participation of developing-country scientists helps keep the panel focused on the concerns of the developing world, but what we really need are more and better research universities,” he continues.

Most South American climate researchers have



Colin Prentice has become aware of issues he'd never considered controversial.

received part of their education in either the United States or Britain, Marengo explains. Their experience abroad can range from a full PhD and postdoctoral project to a two-month stint in a research lab. Those in Commonwealth countries, who already know English, have a head start towards an international career.

Collaborations between industrialized countries and the rest of the world help to establish a network. “We are trying to involve the developing world in our climate-modelling efforts,” says John Mitchell of the Hadley Centre for Climate Prediction and Research in Bracknell, UK. “For example, we have developed a regional model that can be run over any part of the globe, and we also deliver the necessary forcing data. We entertain collaborations with a group of West Indian countries, and the model is already being run over India, China and South Africa.”

The importance of local efforts to invest in monitoring and researching the environment is becoming more obvious to governments all around the world. The IPCC does not directly provide financial support for research. But although it is hard to say what would have become of climate research without the IPCC, the raised profile of climate research (and the appealing possibility of being cited in an IPCC report) has probably stimulated funding agencies in industrialized and developing countries alike.

Joyce Penner, professor of atmospheric, oceanic and space sciences at the University of Michigan and lead author of the chapter on aerosols, points to the cultural difficulties that can arise through international collaborations. “At one point, there was a heated discussion between the American authors of three

Getting up to speed

Besides providing governments with the best available data on climate issues, the Assessment Report by the Intergovernmental Panel of Climate Change (IPCC) can also be used as a primer for scientists wanting to get up

to speed in the field.

Gerald Meehl of the National Center for Atmospheric Research in Boulder, Colorado, who is a lead author on the chapter on projections of future climate change, praises the quality of

the report on the scientific basis of climate change, which costs £34.95 (US\$50). Scientists living in developing countries are eligible to receive a free copy of the report, but those working in industrialized countries have to pay for it.

“We hired a new scientist for our group recently,” says Meehl, “and we just asked him to read a chapter a day.

“Within a short time he’ll be familiar with the current state of human knowledge of climate science.”

H.L.

different chapters,” she says. “A contributing author from Africa later told me that he felt extremely uncomfortable with the aggressive tone of the argument, which we had not fully realized. I think we can all learn from these interactions.”

EASY ACCESS

Those who are involved in the IPCC find it immensely beneficial to their work. Colin Prentice, a director of the Max Planck Institute for Biogeochemistry in Jena, Germany, says: “I would describe the work on the report as a crystallization of knowledge. I was made aware of controversies over issues that I had never thought about as controversial.” He found it “absolutely worth it” to be the lead author of the chapter on the carbon cycle.

Penner also finds the effort “enormously good for the science” and thinks that many people of different interests, from policy-makers to interested lay people, can benefit. “The IPCC reports make the science available for everybody,” she says.

The reports from the three working groups are thought to be fair assessments of the uncertainties and unresolved issues as much as of agreed knowledge. “I never felt any pressure from above to reach consensus, prematurely or otherwise,” says Pierrehumbert. “There is room to consider many viewpoints,” he explains.

Prentice adds that the review process will pick up anything that might have been missed in the writing of any chapter. Those involved are convinced that the reports provide a true representation of the state of knowledge, as far as that is possible — even though this does not hold quite so true for the shorter summaries written for policy-makers. “Policy is unfortunately driven by oversimplification, as policy-makers need crisp statements to think about,” says Pierrehumbert.

INTO THE FUTURE

Yet it is the political relevance of the topic that has kept the IPCC going over the years. Without the funding (at least for travel and some technical support for developing countries), and the policy-makers’ need

Change at the helm

Rajendra Pachauri is the new chairman of the Intergovernmental Panel on Climate Change (IPCC), elected in April together with the leading committee of 30 co-chairs and vice-chairs of the working groups and task force. The replacement of the American atmospheric scientist Robert Watson by Pachauri, whose background is in economics and industrial engineering, caused some stir.

The new head is not expected to exert much influence on the wording of the reports. “External influence from either the working-group chair or IPCC chair was really

not a factor. I expect this process will continue with the new chair,” says Gerald Meehl, one of the current report’s authors.

The chairman’s role is to supervise the panel’s overall agenda and orchestrate the process of the assessment as a whole. Watson also took responsibility for presenting the panel’s findings to policy-makers. The new chairman will decide how best to provide a scientific basis for the negotiations beginning in 2005, before the next IPCC report, for new targets on greenhouse-gas emissions in the context of the Kyoto Protocol.

H.L.

for information, scientists would probably not be able to summon the collective energy necessary to produce the assessment reports.

Prentice says that he would not want to be an IPCC lead author more than once, and Penner generally agrees. “I would have to do it differently, ask other people to contribute more,” she says. “I essentially worked every weekend for two years, and it took me a long time to recover from the strain.”

One way to relieve the demands on the scientists involved is to extend the assessment periods. Initially, the target period between reports was five years, but the third assessment report was published a year late — not surprisingly, considering the enormous amount of work that had to be done on surveying the literature, writing and reviewing. The target periods have now been revised to six years, and some say even ten years would be sufficient.

There is no lack of good reasons to carry on. Many scientific issues are not at all resolved, one of the important examples in Penner’s realm being the indirect aerosol effect, which exerts a cooling influence on climate by increasing cloud cover, though little is yet known about its magnitude. Politicians need scientific input more than ever.

“There was a time in the mid-1990s when we felt we had delivered the facts about climate change to the politicians and thought the ball was in their court,” says Meehl. “But as the difficulties with the ratification of the Kyoto Protocol have become apparent, the question of adaptation, instead of mitigation, is becoming more and more important.”

It seems that climate research will be on the agenda for quite some time.

Heike Langenberg is a physical sciences editor at *Nature*.

International effort: Japanese students at the International Pacific Research Center in Hawaii.



Chance to make a difference in Japan



Vast power: the Earth Simulator.

Japan has the computer power for climate-change research — it just needs to attract people to use it, says Robert Triendl. Solving the skills shortage may involve a long-term change in strategy.

In the early years after the Second World War, young meteorologists in Japan had little choice but to go abroad; positions were few, equipment primitive and funding scant. During the 1950s and 1960s, many went to work in the United States. Syukuro Manabe, a pioneer in climate modelling and research on global climate change, was one of them.

His return in 1997 may be a sign that things are changing. Although investment in human resources and software development lags behind spending on infrastructure, the amount poured into computational facilities is proving a draw. It may even solve other underinvestment problems, by attracting the best and the brightest to Japan to work in fields such as climate-model development.

Manabe returned to head the Global Warming Research Program at the newly established Frontier Research System for Global Change. Its facilities on the outskirts of Yokohama are an integral part of a long-term strategy for research into regional and global climate change promoted by the Japanese government.

Another part of Japan's climate-research strategy is the Earth Simulator, a computer system dedicated to research in the geosciences (see *Nature* **416**, 579–580; 2002) at least an order of magnitude more powerful than any other system available to climate scientists.

DEVELOPING HUMAN RESOURCES

Both facilities have had some investment in human resources. As well as encouraging international links (see 'An international angle', right), the Frontier system offers 160 limited-term positions — just enough to absorb a stream of new graduates coming out of newly established PhD programmes at large national universities, points out Toshio Yamagata, a professor in the department of Earth and planetary science at the University of Tokyo.

But although the Frontier system and other priority funding schemes — such as the CREST (Core Research

for Evolutional Science and Technology) programme administered by the Japan Science and Technology Corporation — are providing new positions and much-needed research money, some argue that a longer-term research strategy is needed. According to Tetsuzo Yasunari, a professor in the department of Earth sciences at Tsukuba University, one important issue for climate research in Japan is the shortage of human resources.

Limited senior-research capacity means that most director positions at the Frontier system have been filled by university professors on joint appointments. Also, says Yasunari, although limited-term positions have surged as a result of the Frontier programme, there has been hardly any change in the number of tenured faculty positions. In Japan's inflexible labour market, this means that many young scientists will find themselves without work once their Frontier contracts expire — and, like several decades ago, they will have little option other than to go abroad.

Although some areas of geoscience — notably earthquake research and volcanology — have received generous funding for over 30 years, the movement of cash into climate research has been relatively recent. Other disciplines, such as palaeontology and geology, have seen hardly any increase in funds in recent years. "We have clearly been lucky," says Yasunari. "Now we need to demonstrate that the funding increases were justified."

Robert Triendl is a freelance writer based in Tokyo.



Tetsuzo Yasunari is eager to prove that funding was justified.

An international angle

Since it was established in 1997, Japan's Frontier Research System for Global Change has hosted scientists from other Asian countries. It also provides funding for the International Arctic Research Center at the University of Alaska in Fairbanks, and the International Pacific Research Center (IPRC) at the University of Hawaii at Manoa. These

were set up under the 'common agenda' signed by Japan and the United States in 1997. "To spend some time with the IPRC is a unique opportunity for Japanese students and young scientists to experience an international research environment," says Toshio Yamagata, a programme director at the Frontier system. R.T.